Can Police-Social Worker Partnerships Improve Crisis Response? Evidence from Columbus, Ohio*

Leah Bevis† Shinjini Pandey‡ Jennifer Hefner§ Steve David¶
Pejmon Noghrehchi‖ Matt Harris**

October 28, 2026

Abstract

Police-clinician co-response is expanding for behavioral crises, but evidence of impact is limited. We evaluate Columbus, Ohio's Mobile Crisis Response program—five citywide police–social worker teams that, due to binding capacity, served only 7–8% of eligible mental health 911 calls. Leveraging random variation in response capacity, we estimate LATEs that describe impacts on healthcare linkage, downside risks of violence or incarceration, 911 system utilization, and direct costs to the city. Co-response roughly quadruples the time responders spend with individuals in crisis, suggesting an improved experience of first response, but it boosts voluntary healthcare contact only under certain conditions. Overall effects on arrests, jail transports, and high-level force are null, but co-response raises involuntary transports by 21 percentage points, an increase concentrated among medium-severity mental health calls, which also see greater use of force. Because co-responders add to rather than replace patrol officers, the program roughly quintuples responder salary costs per call. No significant impact is observed on repeat 911 engagement, though suggestive evidence points to both improved crisis resolution and increased trust in 911.

JEL Classification: K42, I10, J22

*We are grateful to the Robert Wood Johnson Foundation for supporting this study and to the Columbus Division of Police for providing the data and essential domain expertise. We also thank Alex Hollingsworth and seminar participants for their helpful comments. This study was approved by the Ohio State University IRB (2021B0195)

†Ohio State (bevis.16@osu.edu)
‡Duke (shinjini.pandey@duke.edu
§Ohio State (hefner.48@osu.edu)
¶Ohio State (david.54@osu.edu)
‖Ohio State (noghrehchi.1@osu.edu)
**Former Sgt at Columbus Police Division

1 Introduction

Police are the default first responders to mental health and substance use related (jointly, "behavioral health") 911 calls in the United States, and in recent years the public has questioned that response model. Both non-police response programs (e.g., Denver's STAR, Albuquerque's Behavioral Health Responders, or Oakland's MACRO) and co-response programs (e.g., Albuquerque's Mobile Crisis Team, Durham's HEART co-response team, or the Columbus MCR teams) are being piloted nationwide as alternative forms of crisis response. Co-responder programs are particularly common; a recent survey of 568 law enforcement agencies found that 41% had a co-responder program, including many in large cities (Uding *et al.*, 2025). These programs can be seen as a "compromise" solution between traditional response and non-police response: their adoption is often more politically and logistically feasible, in part because institutional stakeholders—particularly police—tend to resist reforms that fully remove law enforcement from the scene (Ba *et al.*, 2024). It is worth asking, therefore, whether co-response programs that build on traditional police response can effectively improve outcomes for behavioral health crises, or whether a more stark break from any type of police response is needed.

We evaluate the impact of a police co-response program in Columbus, Ohio – the Mobile Crisis Response (MCR) program – wherein Columbus police partner with social workers, or clinicians, to respond to primarily mental health and suicide related crises. Using quasi-experimental variation in MCR team capacity and therefore likelihood of response, we use police dispatch data from 2018-2024 to evaluate the causal impact of receiving co-response rather than traditional police response. Our instrumental-variables strategy yields a Local Average Treatment Effect (LATE) for medium- and high-priority mental health and suicide calls received during MCR operating hours (10 a.m.–midnight). Like the rest of the country, behavioral health challenges are a growing concern in Columbus (where demographics generally track the U.S. as a whole): a 2024 assessment found that over half of the local county's residents needed behavioral health services but more than 40% went without care (Alcohol, Drug and Mental Health, 2024). Also tracking nationwide trends, concerns about police violence particularly as pertains to Black residents and those experiencing behavioral health crises were elevated during our sample period. In the year before and during our sample period, three Black men were killed by Columbus police after their family called 911 during a mental health crisis.

Despite a rise in co-responder programs for behavioral health, little is known of their impact in improving call outcomes. Most quantitative evaluations are non-causal, small-sample studies focused on program perceptions rather than incident outcomes such as arrest, jail, use of force, linkage to care, involuntary holds, or subsequent health/treatment status (Shapiro *et al.*, 2015; Watson *et al.*, 2019; Marcus and Stergiopoulos, 2022). One very small randomized control trial in Indianapolis found that co-response did not change rates of EMS transport, jailing, outpatient health contact, or emergency department visits, relative to traditional police (Lowder *et al.*, 2024). Yet the study was under-powered with only 686 behavioral health calls randomized and poor compliance. Two even smaller RCTs evaluated a more limited set of outcomes, finding no impact on 911 call volume in one case and a slight reduction in involuntary hospitalization in another (Yang *et al.*, 2024; Yang and Lu, 2024). A difference in difference study of a co-responder program in San Mateo County, California concludes that the program had little impact

on eligible call volumes or arrests, but decreased involuntary transports by 16.5 percent (Dee and Pyne, 2026). However, pre-trend tests reject parallel trends for three of the four primary outcomes, and involuntary transports fall at least as much during hours when no clinicians are on duty as they do overall, a pattern inconsistent with a program effect.

Most similar to our paper, Ba *et al.* (2025) use a differences in differences framework to find that while Durham's HEART program reduces crime reports and arrests, this is driven by the purely civilian responder teams (non-police response). The HEART program also deploys police-clinician co-responder teams, and they reported crimes and performed arrests at similar rates as traditional police. The largest reductions in crime and arrest reporting were for nonviolent, "quality of life" charges, suggesting that HEART's civilian teams divert low-risk calls, often around homelessness, from the criminal legal system. Use of force, often occurring at the highest-risk incidents, was unaffected by HEART response. However, the program may nonetheless build trust in the 911 system, as follow-up calls increase in the months following HEART response, driven by increased reporting of less-severe incidents.

Relatedly, Davis *et al.* (2025) study the non-police CAHOOTS program in Eugene, Oregon, using two complementary research designs. They find that expansions of CAHOOTS coverage to new hours and areas reduced arrests (especially those arising from mental health hold authority), increased access to medical services (EMS transports), and reduced follow-up 911 calls in the following weeks. Noteworthy for our work, the reduction in arrests comes primarily from CAHOOTS teams co-responding with police rather than responding alone. When CAHOOTS responses are instead driven by short-term variation in team availability, they mostly reach calls that would otherwise have received no response at all, and program impacts are smaller. The measurable impact of adding CAHOOTS teams to police response provides a juxtaposition to the null impact of the co-response teams studied by Ba *et al.* (2025).

Co-responder programs might in theory generate benefits at both the individual and city level. For individuals in crisis, there are three potential pathways. One, the presence of a clinician may improve healthcare provision in the moment. While clinician-individual interactions are not seen in the data, one indicator for quality of care is the time spent on scene, which facilitates assessment, de-escalation, safety planning, and coordination with or referrals to health and social services. Two, clinician-officer pairs may connect individuals to more appropriate treatment facilities, e.g., avoiding hospitals without psychiatric units, or facilitating use of free resources for individuals without high-quality insurance. Three, the program's emphasis on de-escalation and non-coercive interaction may reduce use of force, arrest, or jailing; clinicians may also help facilitate voluntary medical transport rather than utilizing involuntary transport in challenging circumstances. At the city level, co-response programs could mechanically increase or decrease direct response costs depending on clinician and officer salaries and on policies regarding the substitution of clinicians for officers. Medium-run costs could also rise or fall, depending on whether improved crisis response builds trust in the system (thus increasing future call volumes) and/or improves crisis-resolution and long-run health (possibly reducing future call volumes and/or hospitalizations). Co-response programs could reduce jailings, too, immediately or in the future.

Given this framework, we estimate the causal impact of MCR using a two-stage least

squares identification strategy, drawing on detailed police administrative data compiled from multiple non-public sources. We instrument for MCR response with the number of mental illness and suicide related dispatches in other precincts of the city in the 50 minutes before dispatch, conditional on dispatch time. Because MCR capacity is highly constrained, surges in MCR eligible call volume reduce the likelihood that the MCR team will be available for a new dispatch. Balance and placebo tests support the instrument's validity, estimates are robust to adding flexible time fixed effects (Blandhol *et al.*, 2026), and the joint test of strict exclusion and monotonicity proposed by Frandsen *et al.* (2023), which several recent applications reject for their main outcomes, does not reject for any of our main outcomes except use of force. A key strength of our data is the presence of text-based call narratives, which allow us to precisely identify MCR-eligible incidents and to measure severity (psychiatric acuity and risk of harm), enabling heterogeneity analyses absent from prior work. The data also cover a broad range of understudied outcomes, for instance, responder time spent on crises, granular medical transport decisions, involuntary transports, repeat calls tracked via de-identified phone IDs, future calls from neighbors, and lower-level use of force.

At the individual-level, we find that on average MCR response meaningfully impacts only two outcomes: it roughly quadruples the time spent at dispatch, and it raises the likelihood of involuntary transport by 21 percentage points. It has no significant effect on the likelihood of voluntary transport or being left at home, though it modestly increases voluntary transports to private psychiatric hospitals. It has no average impact on the probability of arrest, jailing or use of high-level force. Because MCR capacity is so constrained (reaching only 7.5 percent of eligible sample calls), compliers are the types of calls that are triaged towards MCR response — medium- and high-priority calls that come in during MCR work-hours and seem to need specialized response.

The average effects of MCR response mask substantial heterogeneity. MCR response increases involuntary transports primarily for mental health related calls (related to depression, psychosis, mania, etc.), rather than for calls regarding suicidal ideation or suicide attempts (which see an insignificant increase in voluntary transports under MCR response). MCR response also increases the likelihood of force being used on mental health calls by 4.6 percentage points. Both impacts seem to be driven primarily by medium-severity mental health calls, for which MCR response substantially reduces voluntary transport and increases both involuntary transport and use of force, though these subsample estimates are imprecise. Use of force may be driven by the involuntary transport itself, which involves moving a person against their will into the MCR police car and then into a hospital for assessment. While these instances are generally traumatic—people describe police involvement in involuntary transport as criminalizing and frightening (Jones *et al.*, 2022; Akther *et al.*, 2019), and involuntary hospitalization can erode trust and deter later help-seeking (Jones *et al.*, 2021)—they can be life-saving, and are legal under Ohio law when a person poses an immediate threat to themselves or others. Judging them normatively is therefore difficult, but evidence from one U.S. setting suggests that for marginal cases, where physicians might disagree, involuntary hospitalization increases recidivism and mortality (Emanuel *et al.*, 2025); that MCR response raises involuntary transports most for medium-severity rather than high-severity mental health calls therefore leaves room for concern.

MCR program effects also depend on implementation and context. The LATE of MCR

response on involuntary transport rose by more than half between the early and later phases of the program (though the difference is not statistically significant), and holding complier composition constant widens rather than closes this gap. The change coincides with the move from third-party to city-employed clinicians, but could also reflect changes in post-Covid mental health or any number of other contextual factors. The average effect of MCR on transport to private psychiatric hospitals appears only in 2022 and 2023, when hospital and Crisis Center closures had left public hospitals overwhelmed with psychiatric cases. Neighborhood context matters too: in higher-income neighborhoods, MCR response dramatically increases time spent at dispatch and voluntary transport (by 34 percentage points), adding medical transports, whereas in low-income neighborhoods, point estimates suggest it substitutes involuntary for voluntary transports. MCR teams also respond more often in higher-income neighborhoods (to 8.2% of mental illness dispatches, roughly 1 percentage point more than in low-income neighborhoods).

At the city level, we find that the MCR program raises direct operational costs, both by increasing the number of responders and the number of responder-hours. Neither MCR clinicians nor MCR officers are treated as substitutes for patrol officers, with whom they generally co-respond. Patrol officer numbers and hours on a call show no significant change with MCR response, while the number of responders increases by 1.9 (from an average 2.6 in non-MCR sample calls), and both responder-hours and responder salary costs roughly quintuple, due to the length of time MCR teams spend with the individual in crisis. For the calls MCR teams are triaged toward, this implies roughly $135 in additional responder salary per call, or roughly $234,000 per year across the approximately 1,700 MCR responses to 911 and non-emergency phone calls annually, though these estimates are imprecise and apply to compliers rather than all mental illness calls.

We also find that MCR response has no statistically significant effect on repeat behavioral health calls overall. This null result could reflect no effect, or two channels working in opposite directions: building trust in 911, which raises future calls, and improving crisis resolution, which reduces future health needs. Several pieces of suggestive evidence are consistent with both channels being at play. MCR response has a negative (though insignificant) effect on the 30-day likelihood of repeat behavioral health calls, consistent with crisis resolution, while it (insignificantly) increases repeat calls unrelated to behavioral health and (significantly) increases calls from immediate neighbors within 180 days, consistent with individual-level and community-level trust building. Finally, MCR response does not increase the number of repeat behavioral health calls or the likelihood of a repeat call accompanied by hospitalization over 180 days, suggesting that any increase in the likelihood of at least one repeat call is not driven by worsening crises. This is consistent with Ba *et al.* (2025), who find that Durham's HEART responses do not change the likelihood of a repeat 911 call within 31 days but are associated with more new calls after 31 days, which they interpret, along with a decline in call priority, as evidence of increased trust in 911.

Our study contributes to the literature on policing and public safety. Much of the economic literature has examined how policing influences public safety through crime reduction, focusing on the role of police force size (Mello, 2019; Evans and Owens, 2007), police presence (Weisburd, 2021), workplace risk (Cho *et al.*, 2024), military aid (Bove and Gavrilova, 2017), and female officer representation (Miller and Segal, 2019). More recently, a growing body of work has shifted toward examining factors that improve

public safety by reducing harm during police-civilian interactions. Existing studies have looked at the role of peers in reducing arrests (Rivera, 2025) and the impact of police field training on officer use of force (Adger *et al.*, 2022). Directly related to our setting, recent evaluations of police co-response and non-police response models in behavioral health contexts by Ba *et al.* (2025) and Davis *et al.* (2025) have found positive impacts on various dispatch outcomes. Our study differs from theirs in its focus and findings. Ba *et al.* (2025) study co-responder teams as one component of a program that also deploys civilian teams, finding that co-response yields outcomes similar to police response for crime reports and arrests; we evaluate a stand-alone police–clinician program and examine healthcare linkage, including voluntary and involuntary transport, alongside outcomes like arrest, jailing, and use of force. And while Davis *et al.* (2025) find that CAHOOTS responses sharply reduce involuntary mental health holds, we find that MCR response increases involuntary transports, particularly for medium-severity mental health calls.

Our study also contributes to literature on the cost-effectiveness of public safety policies. One group of papers has quantified the fiscal costs that policing practices and reforms impose on city and police budgets, such as settlement payments from misconduct and civil rights lawsuits against officers (Rozema and Schanzenbach, 2019; Schwartz, 2014). A second group of papers has attempted to assess the welfare impacts of public safety policies by going beyond program and labor costs to quantify their social benefits. Recent studies have valued the crime reductions from police hiring (Weisburst, 2019), the averted fatal and non-fatal use of force from body-worn camera use (Williams Jr *et al.*, 2021), and the arrests and crime reductions from police and non-police crisis response programs (Ba *et al.*, 2025; Davis *et al.*, 2025). Our study contributes to this literature by documenting a case in which alternative responders do not substitute for police: whereas CAHOOTS teams often replace police responses (Davis *et al.*, 2025), and labor cost estimates in Durham assume that co-responders replace a police response and thus raise costs by only $12 per call (Ba *et al.*, 2025), we observe all responders on each call and find that MCR co-response adds roughly $135 per call, in part because department policy does not count clinicians toward the two responders most of these calls require. The benefits are harder to value: co-responders spend far more time with individuals in crisis, which crisis-response guidance and interviews link to better care, and repeat call patterns suggest improved crisis resolution and increased trust in 911. The increase in involuntary transport is harder still to value; it may represent either necessary intervention or harmful coercion, both with long-term cost implications (Emanuel *et al.*, 2025).

The remainder of the paper is organized as follows: Section 2 describes the MCR program in Columbus, our data, and summary statistics. Section 3 presents our identification strategy. Sections 4 and 5 discuss the impacts of MCR response on individual- and city-level outcomes, respectively. Section 6 concludes with a discussion of our findings and their policy implications.

2 Institutional Context and Data

2.1 Columbus Co-Responder Program

Columbus piloted the Mobile Crisis Response program in June 2018. The program pairs 5 police officers trained in de-escalation and crisis intervention with 5 mental health clinicians to jointly respond to mental health and suicide related crises. Of the five established officer-clinician teams, 1-3 are on call between 10am and midnight all days of the week (shift schedules shown in Figure A1). These teams respond to calls city-wide, unlike patrol officers who respond almost exclusively within their assigned precinct.[1] During our study period, MCR teams jointly responded to an average 5.6 dispatches citywide per day, representing only 7.5% of the average of 64 mental health or suicide related dispatches occurring per day. This constrained capacity is key to our identification strategy.

Throughout our sample period, any officer could apply for an MCR position, and union hiring rules required that positions be filled by seniority. Thus, while many applicants were surely motivated by interest in mental illness, neither this interest nor any relevant skills determined who was hired from the applicant pool. Initially, in fact, the MCR openings were considered "easy" assignments, a break from patrol. Anecdotally, this led to a mix of officers, some intrinsically motivated to serve people with mental illness and others not.

Calls are dispatched to MCR teams through the same system that routes all emergency calls in the city. Initial "call-takers" at the Emergency Communications Center sort 911 and non-emergency calls into those that go to the Division of Fire (including all physical health-related calls) and those that go to the Police Department (including all behavioral health calls). Police calls are then transferred to "dispatchers," who assign one of roughly 130 police codes; each code carries implications for priority level (1–5) and the number of officers dispatched, though both can be adjusted manually.[2] MCR teams primarily respond to calls coded as "mental disturbance" (49% of MCR dispatches) or "suicide attempt" (21%), and also to mental health and suicide-related calls coded otherwise, e.g., "wellbeing check" (7%), "disturbance" (4%), or even "person with a gun", "person with a knife", or "domestic violence" (each 2%). Many of these calls also involve some element of substance use, addiction, and/or homelessness. Most of the codes MCR teams respond to require at least two officers by CPD policy, with more at the dispatcher's discretion, so the majority of MCR dispatches involve one or more police officers alongside the MCR team. In some cases those officers arrive first; in others, the MCR team does.

MCR teams are dispatched to an incident in one of four ways. First, the dispatcher may dispatch an MCR team directly to a call as it comes in, if an MCR team happens to be free or will be free soon. Second, police may radio MCR teams to request back-up

[1] For patrol purposes, the Columbus Division of Police has divided the city into five police zones. Each zone is split into four precincts (precincts 1-20), and each precinct is further split into several cruiser districts. Starting from April 30, 2023, the patrol zones were restructured and an additional zone six was added. The number of precincts remains unchanged – but did boundaries change??.

[2] The police code decision follows a rubric but leaves dispatchers significant latitude to make their own judgment calls (Mission Critical Partners, 2025).

at an incident. Third, MCR teams may finish the paperwork for an incident, scan the Computer-Assisted Dispatch (CAD) terminal for recent, high-priority behavioral health calls, and self-dispatch to the incident that seems most pressing, volatile or violent based on dispatcher remarks. Fourth, behavioral health professionals (social workers, clinicians, librarians) or even families with previous MCR engagement may call the MCR teams directly on their work mobiles; the MCR team then calls into dispatch themselves to create an "officer-initiated" dispatch.[3] In all four cases, MCR teams are triaged by somebody—the dispatcher, another officer, themselves, the requesting party—towards certain types of mental health calls. Thus, the "mental disturbances" and/or "suicide attempts" the police and MCR teams respond to will not be directly comparable.

Both patrol officers and MCR teams transport individuals in serious crisis to health facilities for in-patient services. While 21% of police dispatches for mental health crises end in such transport, 41% of MCR dispatches do, reflecting the increased acuity of the incidents that MCR teams respond to. Three options exist for in-patient services: a free crisis center, private psychiatric hospitals, and public hospitals. The one free crisis center in Columbus, operated by Netcare Access, iteratively restricted it's willingness to take police drops over the course of our sample period, and essentially closed in 2023.[4] Private psychiatric hospitals abound, but police rarely drop individuals to them due to the facilities' reluctance to admit uninsured or Medicaid patients and most patients' own avoidance of these places due to high treatment costs. MCR officers and clinicians also consider the private psychiatric hospitals to provide lower-quality care than public hospitals, a perception in line with multiple investigations into patient abuse in private behavioral health hospitals in Ohio.[5] It is possible that some patrol officers share this perception. In any case, for all these reasons, the vast majority of individuals who need immediate, in-patient care (just over 90% of the nearly 35,000 health facility transports in our dataset) are dropped to public hospitals.

However, during our sample period, initially only three and later two public hospitals had psychiatric units. (The primary children's hospital also has a youth psychiatric unit.) Admittance to a hospital without a such a unit can result in poor care, as individuals may be held with minimal treatment until transferred to an appropriate facility, or treated by a doctor with limited knowledge of their condition. In theory, MCR teams and many police officers know this. However, both patrol and MCR officers report that hospital locations shape their decisions regarding where to transport an individual in crisis. All in-patient behavioral healthcare facilities are mapped in Figure A2.

In cases where an individual represents a "substantial risk of physical harm to self or others", police officers may transport them involuntarily – without their consent

[3]Standard officers also initiate dispatches, often in response to incidents observed on patrol. For analysis we drop all officer-initiated calls, by officers or MCR teams, analyzing only public requests for assistance.

[4]Netcare began restricting the types of police transports it would accept towards the end of 2019, and closed completely on November 25, 2023. While in theory the center opened under Recovering Innovations International in early 2023, the ability for police to drop those in crisis remained so restrictive that it can be almost discounted as an option for 2024.

[5]A 2019 investigation into Sequel Pomegranate, a Columbus psychiatric facility for teens, found dozens of reports of abuse and sexual assault, many substantiated. Sequel Pomegranate now operates as Torii Behavioral Health. MCR police officer Matt Harris (a co-author on this paper) took the extraordinary step of warning Ohio regulators about the physical and sexual assaults he routinely observed at Mt. Carmel Behavioral Health psychiatric hospital. Mt Carmel Behavioral has since shut down.

– to a healthcare facility. This is legal under Ohio Revised Code Section 5122.10; psychiatrists, licensed clinical psychologists, licensed physicians, and a few other credentialed professionals may do the same. Ohio Revised Code Section 5122.01(B) sets out five conditions, four of which specify the manifest evidence required to independently establish this "substantial risk". The fifth condition — concerning a person unlikely to survive safely without help, and with a history of non-compliance with treatment — cannot by itself constitute legal grounds for involuntary transport. Once involuntarily transported to a hospital or the crisis center, a medical provider decides on inpatient mental health intervention – thus, involuntary transport does not necessarily equate to involuntary hospitalization. During our data period, involuntary transports were common: among nearly 34,000 healthcare transports, 28% were involuntary. While a quarter of non-MCR transports were involuntary, 40% of MCR transports were.

Involuntary transports were also a major source of tension between clinicians and police in the program's initial phase, ultimately reshaping clinician involvement. Initial clinicians were employed through Netcare Access, a third-party behavioral healthcare facility that also ran the free crisis center. Appointed as "health officers", these clinicians held the legal power to involuntarily transport—but within a team, officers and clinicians frequently disagreed over its appropriateness. Officers felt clinicians sought to transport individuals posing no immediate risk (deprivation of civil liberties), while clinicians felt officers underestimated risk for lack of specialized training (deprivation of healthcare). In late 2019, CPD and Netcare agreed that when an officer disagreed with his or her partner clinician, the individual would be left in place and a request for involuntary transport sent to the Franklin County Probate Court.[6,7] This arrangement lasted through December 31, 2020, when the CPD-Netcare partnership ended, largely due to this ongoing tension. From January 1 to July 1, 2021, MCR officers responded alone while Columbus Public Health hired and trained replacement clinicians; by around July 1, all five had joined their teams.[8] Crucially, however, the new clinicians were not granted "health officer" status, leaving MCR officers the sole legal decision-makers on involuntary commitments.

Thus, two important factors varied across the early and late stages of the MCR program. The first is the clinicians: third party Netcare clinicians staffed MCR through 2020, while city-employed clinicians were firmly in place by 2022. This change, as well as the Netcare-CPD probate agreement in place for 2020, has potential implications for involuntary transport since Netcare clinicians were "health officers" with the legal power. Second, in 2021 the Columbus Police launched a mental health call diversion program, the Right Response Unit.[9] Piloted in June of 2021, the program was designed to handle low-acuity mental health calls on the phone, rather than by dispatching officers. Right Response clinicians provide callers in crisis with relevant resources, such as referral to a

[6] Any individual may request an involuntary medical transport via Probate Court. The judge there decides whether it is legally merited given the evidence.

[7] MCR officers recall that disagreements often regarded whether a history of non-compliance with treatment and risk of relapse alone warranted involuntary hospitalization; under Ohio Revised Code Section 5122.01(B) it does not.

[8] One Netcare clinician remained, moving to work under Columbus Public Health. The other four were new. Start dates varied slightly across the five clinicians and no record is held of those dates.

[9] Another mental health program existed but did not change during our data period: the Crisis Intervention Team program provides officers with 40 hours of specialized training on mental health, crisis resolution, and de-escalation. It has been mandatory for all new recruits since 2018.

warming shelter or a drug rehabilitation clinic, or connection to an existing case worker. For low-acuity, low-risk situations, this may be all that is needed. In other cases, clinicians dispatch either police offices or an MCR team for in-person response, generally staying on the line until a first responder arrives. The program has expanded steadily: piloted on weekdays from noon to 4pm, hours were extended to 8:30am–4:30pm in January 2022, and to midnight in October 2022.[10] By the end of our data period, Right Response clinicians handled 17% of mental illness related dispatches.

The COVID-19 pandemic provides a further temporal context for both changes. The pandemic's effects on mental health were concentrated in 2020: population-average symptoms of anxiety and depression rose sharply during the lock-down phase but had largely receded toward pre-pandemic levels by 2021, even as elevated symptoms lingered for some subgroups (Sun *et al.*, 2023; Ettman *et al.*, 2022). This transient shock sits atop a slower, national decline in mental health across our full sample period, e.g., with depression prevalence rising from 8.2% in 2013–2014 to 13.1% in 2021–2023 (Brody and Hughes, 2025) and reports of persistent sadness and suicidality among high-school students climbing steadily (Centers for Disease Control and Prevention, 2024). However, if mental health in any single year was differentially impacted by the COVID-19 pandemic, it is 2020—when Netcare clinicians still staffed MCR, and before the Right Response Unit came online. It is possible that both voluntary and involuntary transports of patients in crisis were also constrained by the state of hospitals during this period.

2.2 Data Sources and Sample

Our primary data source is the Columbus Police Department Computer-Assisted Dispatch (CAD) dataset, which holds details on the universe of all 5.1 million dispatches that occurred between January 1 2015 and December 31 2024. This dataset was obtained under a data use agreement. Each observation represents a police dispatch, either responding to a 911 or non-emergency call or initiated by an officer. A dispatch could be one officer, one MCR team, multiple officers, or a mix of MCR teams and patrol officers. Patrol time is not included in police dispatch data; each dispatch represents a unique interaction with the public. For each dispatch we observe a variety of data fields including GPS location, a series of time stamps, the assigned police code, priority level, incident location, disposition, and transport history. Critically, a remarks field contains call-taker notes summarizing information received from the caller (for officer-initiated dispatches, the caller is the officer). A sample view of the data is presented in Table A1.

Our analysis sample contains only dispatches responding to mental illness related calls, the primary domain of MCR response. To identify mental illness dispatches we used both the police code and dispatcher remarks fields. We first identified the police codes assigned to mental illness calls: mental disturbance (10-16B), suicide attempt (10-47A), overdose (10-34O), intoxicated person (10-19), and indecent exposure (10-44B). Yet many mental illness calls are assigned other codes (e.g., as welfare checks, disturbances, domestic violence), and so we used the call remarks field to re-classify dispatches, as

[10] While initially calls were passed to Right Response by dispatchers, in late 2021 an "option 3" was added to 911, letting callers reach Right Response directly for mental health crises when available (reverted to standard dispatchers when clinicians were busy).

explained further in Appendix B. Finally, we keep all mental illness dispatches that related to suicide attempts or general mental health, but discard those primarily related to substance use or abuse (as MCR teams are not triaged towards such calls).

A few other levels of selection lead to our final analysis sample. We drop from the sample all dispatches prior to the advent of the MR program in June 2018. We also restrict our analysis sample to call-based dispatches only. This is because officer-initiated dispatches are not as strongly impacted by our instrument, and because MCR-initiated dispatches are not comparable to patrol-initiated dispatches. For instance, this choice means that we drop all dispatches where MCR officers are responding to calls they get on their personal work phones; only calls coming through the 911 or non-emergency police number are included. Last, we drop MCR follow-up dispatches, where the MCR team arrives more than 1 hour after the first team was dispatched; our goal is to examine the impact of first-response by a police-clinician team, not the impact of police-clinician follow-up.

The ultimate analysis sample is 147,678 observations, each representing a unique incident that one or more police or MCR teams responded to.[11] MCR teams were dispatched to 11,138 of these incidents, which we refer to throughout as "MCR dispatches" or "MCR-handled calls". Of these, 82% also received a police dispatch, so MCR teams usually respond alongside police.[12] For comparison, Davis *et al.* (2025) document that police are involved in 38% of CAHOOTS responses. When co-responding with police, MCR teams arrive first or arrive simultaneously with police 21% of the time.

Using a unique dispatch identifier, data from three other police datasets were merged into the analysis sample. We merged in use of force records for the sample period (obtained by public records requests). These use of force data reflect the database maintained by Columbus Police Internal Affairs Bureau, and they reflect only "serious" use of force: levels 2-7, or levels 0-1 "with injury". However, we also merge data that was exhaustively digitized from paper use of force forms from 2023 through 2025, which includes use of force levels 0-8 (with 8 being fatalities). Second, we merged in arrest reports from a separate police database, Premier1, to identify incidents that resulted in an arrest and to determine the charges filed. We classified charges using NIBRS offense codes, as well as drawing out public-order, drug, or addiction-supporting charges, since these charges are most likely to result from mental illness related dispatches. Third, instances of involuntary transport were merged in from the Premier1 database, which maintains electronic records of aid and transport forms, as well as mental health contact forms completed by officers when they decide to transport an individual involved in a call. Further details on the use-of-force datasets, arrest classification, and the Premier1 data are provided in Appendix B.

We also use the American Community Survey (ACS) five-year estimates to classify all Columbus census tracts into three socio-income categories — lower-income minority, lower-income white, and higher-income. Figure A3 shows the spatial distribution of the three neighborhood categories in Columbus. Panel (a) shows neighborhood categories

[11] This includes 61,842 observations that were not explicitly coded by dispatchers as mental illness related – "suicide attempt" or "disturbance / mental".

[12] MCR teams are more likely to respond alone when self-dispatching, and thus the co-response rate is greater in our sample, which includes only call-based dispatches, than it is for all MCR dispatches. Within our sample, co-response with police rises over time, from 61% of MCR dispatches in 2018 to 90% of MCR dispatches in 2024.

based on the 2020-2024 ACS five-year estimates, which were used for dispatch data from 2020-2024, whereas panel (b) shows neighborhood categories based on the 2015-2019 ACS five-year estimates, which were used for dispatch data from 2018-2019.[13] Notably, most of the lower-income, minority census tracts are on the east side of Columbus. Two neighborhoods, Hilltop in the south-west and Linden in the north provide most of the lower-income white census tracts. Thirty, twenty-six and forty-four percent of dispatches related to suicide and mental illness are observed in lower-income minority, lower-income white, and higher-income areas, respectively, which almost perfectly reflects the spatial distribution of dispatches more generally.

2.3 Time Spent as a Proxy for Care Quality

We use time spent at dispatch as a proxy for the quality of care that first responder teams provide, for two reasons. First, practitioner guidance and training for mental health crisis response emphasize that high-quality crisis response takes time. To qualify for enhanced federal Medicaid funding, mobile crisis services must provide "screening and assessment, stabilization and de-escalation, and coordination with and referrals to health, social and other services and supports," and teams may also need to initiate safety planning (Centers for Medicare & Medicaid Services, 2021), which commonly takes 20 minutes or more (Stanley and Brown, 2012; Boudreaux *et al.*, 2017). Emergency psychiatry's consensus guidelines on verbal de-escalation conclude that "[e]ven the most complicated cases can be managed with a little additional time" (Richmond *et al.*, 2012). The Police Executive Research Forum's ICAT 2024 training, designed for encounters with mental health or situational crises, instructs officers to "slow the situation down" because "in most situations, time is on our side".[14] Crisis Intervention Team training materials from Columbus, Ohio similarly advise officers to "give the situation time".[15] Related economics evidence finds that the quality of physician care depends in part on time spent with patients (Chan, 2018; Silver, 2021; Freedman *et al.*, 2021; Neprash *et al.*, 2023).

Second, in a linked qualitative project described by Hefner *et al.* (2026), Columbus residents who had received an MCR response noted that the extra time spent on scene was notable and appreciated, and led to real safety differences. This theme was particularly prevalent among interviewees with a long history of mental health crisis calls, and thus a basis for comparison with traditional police response. For instance, one interviewee

[13] GPS coordinates in the CAD data allowed us to place each dispatch within a U.S. census tract. Tracts with a non-white population of at least 50% and a median household income at or below the income cutoff were classified as lower-income minority; tracts with a non-white population below 50% and household income at or below the income cutoff were classified as lower-income white; all other tracts were classified as higher-income. The income cutoff – set at 200% of the federal poverty line for a family of four – was $60,000 under the 2020-2024 ACS five-year estimates and $50,000 under the 2015–2019 ACS five-year estimates.

[14] The ICAT guide explicitly challenges the "culture of speed" in policing, asking: "If time is available, why do we rush?" Found in 2026 at: https://legislature.vermont.gov/Documents/2026/Workgroups/House%20Judiciary/Bills/H.613/Witness%20Testimony/H.613~Christopher%20Brickell~ICAT%20Training%20Guide%20February%202024~2-25-2026.pdf.

[15] The same materials note that officers "may need to repeat requests," and that "[y]ou may have to repeat this assuring message many times before the individual can respond to it". Found in 2026 at: https://nisonger.osu.edu/wp-content/uploads/2016/11/CIT-Framework.pdf.

noted that MCR responders *"....come in hoping to build a relationship, almost like build a rapport. They come in and they start chatting about maybe the person's pet, or they make comments...Rather than just responding to the crisis, they tried to build a little relationship before going after that."* Another interviewee connected the extra time to a life or death outcome: *"So, if they [MCR responders] weren't patient with her—because they were here for like an hour and a half, and obviously they have lots of calls and a lot to do—but, if they weren't patient with her that day, I mean, she could be dead for sure."* Qualitative studies of crisis response in other settings similarly find that people in crisis value responders who are calm, empathetic, and respectful (Boscarato *et al.*, 2014; Pope *et al.*, 2023); our interviewees' accounts suggest that time on scene is a key input into creating such an environment, and through it, into both quality of care and safety.

We construct time spent at dispatch as the difference between recorded dispatch time (when the first responding team is dispatched) and closing time (when the last responding team leaves), because arrival time is often missing. While stamps for call time, dispatch time, and closing time are automated, arrival time is recorded by police as they remember to do it; 20% of the time it goes unrecorded by at least one unit, and for 6% of dispatches no police units record it. It may also be recorded inaccurately in many cases. (Thus, time to arrival, constructed as the difference between call time and arrival time, is unavoidably errored, with non-randomly missing values.) The distributions of time spent at dispatch and time to arrival are highly skewed and never zero; we log transform them for analysis.

2.4 Gauging Risk and Psychiatric Acuity

We construct a crisis severity measure from the call-remarks field of the CAD data. Starting with remarks from mental illness calls that dispatchers assigned priority 1 or 2, we manually compiled a dictionary of words and phrases that signal urgent or high-risk situations. To build a more comprehensive set of keywords, we used a large language model (Mistral, run locally through Ollama) to extract candidate terms from 2,500 randomly sampled remarks containing at least 10 words, which we manually reviewed before adding to the dictionary. The final dictionary contains 696 terms and phrases, including common misspellings and variants (e.g., "schizo," "skitzo," "scizo" used as a shorthand for schizophrenia). A complete list of keywords is provided in Appendix B. Each incident's composite severity score is the total number of times dictionary terms appear in its remarks.

We group the terms into four categories: psychiatric acuity, mostly diagnoses and symptoms (e.g., bipolar, hallucinating, homicidal); medical emergency (e.g., seizure, bleeding, weak pulse); non-clinical descriptions of a person in crisis, including lay terms, substance use, and signs of a break from reality (e.g., crazy, lunatic, ranting, demons, "camera is in his eye"); and risk of harm, covering violence or escalation (e.g., threatening, berserk, choking, "putting hands on"). This lets us examine MCR impacts by overall severity and separately by psychiatric acuity and risk of harm, since MCR teams consider themselves most effective on psychiatrically acute calls with a high risk of harm.

Among mental illness calls, the severity score ranges from 0 to 30 and increases with

dispatcher-assigned urgency.[16] We classify calls with a composite score of four or more as high severity, two or three as medium, and zero or one as low, yielding 32.3% high-, 38.4% medium-, and 29.3% low-severity calls. We classify psychiatric acuity as high if remarks contain two or more psychiatric terms, medium if they contain one, and low if they contain none, and we define risk-of-harm levels the same way. That nearly 50% of involuntary transports occur among high-severity calls and nearly 35% among medium-severity calls provides a useful check on our classification, suggesting that the measure captures meaningful differences in call severity.

2.5 Descriptive Statistics

Figures 1 and 2 illustrate time trends in mental illness dispatches over our sample period by providing binned dispatch counts. Figure 1 demonstrates three points. First, mental illness dispatches vary seasonally; this seasonality is greater for general mental health dispatches than for suicide related dispatches. Second, while mental health focused dispatches were on a gradual rise from 2018-2021 (suicide related dispatches were holding steady), both mental health and suicide related dispatches fell slightly in 2022, staying on that new trajectory through 2025. Third, the average severity of these dispatches rose sharply in 2022, and remained high after. While Figure 1 demonstrates this fact using the severity index – a count of all psychiatric acuity-related or risk-related words or phrases – the same trend can be seen in acuity-only word counts, risk-only word counts, or counts of very specific phrasing around, e.g., weapons or schizophrenia (Figure A4).

Both the reduction in mental health dispatches and the increased severity of those dispatches, post-2022, is in line with the goals of and anecdotally-recalled effects of the Right Response Program. By shifting the lower-acuity, lower-risk calls to a telehealth response, need for in-person response should have been diminished. The average acuity and risk of remaining calls therefore also climbs. In Figure 2, we observe that while the severity of mental health dispatches handled by patrol officers rises in 2022, the severity of MCR dispatches rises more. This makes sense, given the triaging of high-acuity, high-risk calls towards MCR. Again, the same trend can be seen in acuity-only counts, risk-only counts, or counts of phrasing around weapons or schizophrenia (Figure A5).[17]

Table 1 provides summary statistics for dispatch outcomes, pooled over our sample period but separating dispatches that received an MCR response from those that did not; Figure A6 plots the corresponding time trends. MCR-handled calls differ markedly across nearly all measures. On average, first responders take longer to arrive (by about 4 minutes) for MCR-handled calls, and then remain on scene far longer (by roughly an hour).[18] An MCR-associated delay of only 4 minutes reflects the fact that another police team often

[16] The mean score is 3.61 for priority one and two calls, 2.32 for priority three calls, and 1.92 for lower priority calls.

[17] Figures 2 and A5 further demonstrate that the risk levels and psychiatric acuity of MCR dispatches were similar to those of police in 2018, but began rising right away. The differential jump in severity of MCR calls in 2022 was due more to a jump in risk of harm than to a jump in psychiatric acuity.

[18] These rates are steady, with a spike in average time to arrival in the summer of 2020 coinciding with protests in Columbus following the murder of George Floyd in May 2020. Columbus police officers were anecdotally observed to pull back from their duties, or "depolice" during these protests, as observed in other cities (Roman *et al.*, 2023).

shows up before the MCR team does.

MCR-handled calls are also more likely to end in transportation to the free crisis center (by 4 percentage points) and to an emergency department with a psychiatric units (by 8 percentage points), though about 1 percentage point less likely to end in transport to an emergency department without a psychiatric unit. These transport patterns shift over time. As the crisis center imposed restrictions on the patients it would accept between 2018 and 2022, transports there fell, especially for MCR dispatches. Perhaps partly in response to that infrastructural gap, MCR transports to emergency departments with psychiatric units rose from about 10% of MCR dispatches at the program's start—similar to the police rate—to about 35% by 2021, staying high though with a dip towards the end of our sample period. Appendix B provides further detail on the construction of transport variables.

On average, MCR dispatches ended in involuntary transport far more often than police dispatches—a 13 percentage point difference (Table 1)—but this gap masks variation over time (Figure A6). Police rates were flat throughout the sample period, with about 5% of dispatches ending in involuntary transport and about 15% in voluntary transport. MCR rates, by contrast, evolved. In the program's first months, just over 5% of MCR dispatches ended in involuntary transport (similar to police), but this rose to 20–30% through 2019 and early 2020, with voluntary and involuntary transports then roughly equal. Then in mid-2020, perhaps due to conflict within the MCR teams over civil liberties, involuntary transport by MCR teams began to fall and voluntary transport to rise. By the late (2022-onwards) phase of the MCR program, about 15% of MCR dispatches ended with an involuntary transport and about 25% with a voluntary transport.

Arrest, jailing, and use of force are all rare in our dataset, though more prevalent for MCR-handled calls. (Appendix B provides more details on the construction of these variables.) When police handle mental illness calls, only 1–2% end in arrest or jailing; not everyone arrested is jailed.[19] MCR-handled calls are about half a percentage point more likely to end in arrest or jailing (Table 1)—a difference that is statistically significant and large relative to the police rate, though small in absolute magnitude. Use of serious physical force is even less common, occurring in 0.18% of sample dispatches handled by police and 0.45% of those handled by MCR teams. Again, the difference is small in magnitude but relatively large: during mental illness dispatches, MCR teams use force at 2.5 times the rate police do. These differences almost surely reflect, at least in part, the underlying selection process for MCR dispatches: MCR units are routed—through self-selection or dispatch triage—to higher-acuity, higher-risk calls. As a result, raw comparisons are uninformative and would fail to capture the causal effect of the MCR program.

[19] One may be arrested but not charged or jailed, arrested and provided with a citation or summons in lieu of jail, arrested and then diverted to a health facility for injuries, etc.

3 Research Design

3.1 Core Specification and Caller Outcomes

We measure the impact of MCR response (vis-a-vis standard police response) via instrumental variables, restricting our analysis to mental illness and suicide related dispatches since these are the domains to which MCR teams regularly respond. Simply comparing the outcomes of MCR and police dispatches would provide a biased estimate of program impact, since MCR teams are triaged towards the most psychiatrically acute and/or highest risk mental health incidents. When estimating Equation 1, we therefore instrument for MCR dispatch to call i in precinct p at time t (MCR_{ipt}) with a conditionally exogenous shifter in MCR teams' capacity to take call i: the number of mental illness and suicide related dispatches, city-wide but excluding the precinct of call i, that occurred in the 50 minutes before the dispatch of call i (Z_{ipt}).

$$Y_{ipt} = \beta MCR_{ipt} + F_{hw} + F_m + \psi_1 t + \psi_2 t^2 + \nu_{ipt} \tag{1}$$

We condition on, or residualize by, daily trends (F_{hw}), annual seasonal patterns (F_m), and an overall time trend ($\psi_1 t + \psi_2 t^2$, where t_i indicate calendar time) so that identification rests only on plausibly exogenous deviations in recent call volume from predictable temporal patterns. Leveraging conditionally random call history, as Davis *et al.* (2025) also do, is not dissimilar to leveraging conditionally random decision-makers like judges, physicians, or attorneys.[20] Adapting from that literature, we might residualize diurnal and annual trends using fixed effects; F_{hw} might contain hour-of-day $\times$ weekend fixed effects, for instance — since diurnal patterns in mental health call volume vary across weekends vs. weekdays but are close to identical within those categories (Figure A7) — and F_m might contain month fixed effects. However, residualization of cyclical patterns can be done more accurately, more smoothly, and with fewer degrees of freedom via tools like cubic splines or truncated Fourier series (CITE, CITE).

We define F_{hw} using the adapted cubic spline displayed in Equation 2. Diurnal trends in dispatch volume are shaped by both patterns in call volume and police shift changes, which occur at either 6am/2pm/10pm or 7am/3pm/11pm, depending on the precinct. Both are evident in the binned means shown in Figure 3. Mental health call volumes dip low in the night, begin rising at the start of the workday, and peak around 8pm. Additionally, in the hour prior to 6am, 2pm, and 10pm, dispatch volumes drop as officers limit new dispatches towards the end of their shift, then rise linearly for 2 hours as a call backlog is swiftly dispatched.[21] Our cubic spline is therefore adapted to enforce the observable linearity between 6-8am, 2-4pm, and 10pm-midnight, while allowing a cubic

[20] See for example Huttenen et al. 2026; Sigstad 2026; Agan, Doleac and Harvey 2023; Garin et al. 2025.
[21] The linear bounce in dispatches is smallest in the early morning when calls are few.

trend in the 3 time windows between those periods.[22]

$$F_{hw} \equiv \sum_{d\in\{0,1\}} \mathbb{1}[W_i = d] \left[\beta_{0d}\, t_i^h + \sum_{k\in\mathcal{K}} \beta_{kd}\, L_k\big(t_i^h\big) + \sum_{w\in\mathcal{W}} \Big(\delta_{wd}\, Q_w\big(t_i^h\big) + \lambda_{wd}\, C_w\big(t_i^h\big) \Big) \right] \tag{2}$$

Here $t_i^h \in [0, 24)$ is the dispatch time of call i in hours, and $W_i \in \{0, 1\}$ indicates a weekend, such that each coefficient is estimated separately by day-type. The linear hinges $L_k(t_i^h) = \max(t_i^h - k,\, 0)$, one per shift-change boundary $k \in \mathcal{K} = \{6, 8, 14, 16, 22\}$, together with the slope $\beta_{0d}\, t_i^h$ form a continuous piecewise-linear backbone. Within each smooth window $w \in \mathcal{W} = \{[0, 6], [8, 14], [16, 22]\}$, spanning $[a, b]$, the terms $Q_{[a,b]}(t_i^h) = \max(t_i^h - a, 0)^2 - \max(t_i^h - b, 0)^2$ and $C_{[a,b]}(t_i^h) = \max(t_i^h - a, 0)^3 - \max(t_i^h - b, 0)^3 - 3(b - a)\max(t_i^h - b, 0)^2$ add quadratic and cubic curvature; the subtracted terms return this curvature to zero at b, so the trend is cubic within each window and linear elsewhere.[23]

We define F_m using the first-order truncated Fourier series displayed in Equation 3. Time t_i^y indicates the within-year time of call i, moving from 0-1 over the course of the year. $\tau_1 = 1$ by definition.[24] Estimated coefficients $\hat{\phi}_1$ and $\hat{\phi}_2$ describe the shape of seasonality. As shown in Figure 4, this Fourier series specification – in combination with the overall trend captured by $\hat{\psi}_1 t + \hat{\psi}_2 t^2$ — does well capturing the (fairly noisy) annual trends in mental health dispatches, namely that they are lowest in winter and highest in summer.

$$F_m \equiv \phi_1 \sin\left(\frac{2\pi t_i^y}{\tau_1}\right) + \phi_2 \cos\left(\frac{2\pi t_i^y}{\tau_1}\right) + \xi_1 t_i + \xi_2 t_i^2 \tag{3}$$

We create Z_{ipt} as a leave-out count of the city-wide mental health dispatches occurring in the 50 minutes before the dispatch of call i. We exclude mental health calls from precinct p in this count to avoids concerns that local call or dispatch volumes may be causing or correlated with local conditions (e.g. a nearby hospital closure, or a major incident on the street that provokes many 911 calls) that may also impact local outcomes. We chose a call history period of 50 minutes because the mean mental health call in our data lasts for 50 minutes, from arrival till closing. However, we show robustness to specifying this period as 30 minutes (the median length of a call), 40 minutes, 60 minutes, or 70 minutes.

To measure the caller-level impact of MCR response, we estimate Equation 1 with outcome Y_{ipt} representing a list of timing outcomes (time to arrival; time spent at dispatch), healthcare contact outcomes (no transportation to a jail or hospital; voluntary transportation to a hospital; specifics on which type of hospital), and downside risk outcomes (arrest; jailing; force used; involuntary transport to a hospital; specifics on which type of hospital) for call i from precinct p at time t.

[22] Enforcing linearity during the 3 shift changes is important, because otherwise the call volume jumps are not well captured. The nature of dispatched calls varies from the low point where primarily urgent calls are being dispatched to the high point where all the low-urgency calls are finally being dealt with.

[23] The trend is continuous throughout, as every term is continuous and equals zero before its first knot.

[24] In a second or third order truncated Fourier series, additional terms are added with $\tau_2 = 2$, $\tau_3 = 3$.

3.2 Heterogeneous Impacts

Our data allows us to examine heterogeneity across dimensions that are important for understanding program impact generalizability. The scope of an alternative response program's mandate is generally a point of contention during the planning phase; in particular, policy-makers are often concerned about sending non-police responders into dangerous situations. We therefore examine heterogeneity by incident type (mental health vs. suicide related) and by call "severity" (tiers of psychiatric acuity and/or risk of harm). How to contract clinicians or other non-police response team members is also a key choice, with some cities employing third party contractors and others creating new spots for city-employed, unionized clinicians. Because Columbus did both, we compare MCR impact across the two eras of clinician employment (2018-2020 under third-party Netcare clinicians vs. 2022-2024 under city-employed clinicians), with 2021 a transition year.

Examining MCR before and after 2021 also makes sense because 2021 is the year in which Right Response was piloted and expanded, and the psychiatric acuity and risk levels of the average mental illness call rose in this year (Figure 2). If the impact of MCR response varies by psychiatric acuity and/or risk levels, we might therefore expect the impact of the program to shift across these two eras due to this change in call composition only. However, for involuntary transports, a second dynamic is at play; we might expect the impact of MCR response on involuntary transports to decline in the post-2022 era for low-severity cases in particular, given that according to police, Netcare clinicians were pushing for involuntary transport in such cases, beyond what police were comfortable or would do on their own. We investigate these dynamics empirically.

Alternative response programs are often promoted within a context of perceived racial inequality in the quality of police response to behavioral health crises. We examine variation in police responses to mental health crises across three groups of neighborhoods—poor, primarily non-White neighborhoods; poor, primarily White neighborhoods; and wealthier neighborhoods—and then ask whether the MCR program's impact widens, narrows, reverses, or leaves those disparities unchanged. (In the context of Columbus, primarily non-White almost always indicates primarily Black.) We also ask whether the MCR program's effect on transports varies with distance to various healthcare facilities—distances that themselves differ across neighborhood groups.

Two heterogeneity tests inform policies around the hiring and training of police officers within alternative response programs. First, we examine heterogeneity by officer experience with the MCR program, to assess the possibility that outcomes change as MCR officers learn. Second, we indirectly test for heterogeneity in program impact across officers by examining heterogeneity in program impact across scheduled shift assignments. Testing directly for heterogeneity across officers is not possible with our instrument; the instrument shifts the availability of MCR officers as a whole rather than of any individual officer, and restricting the sample to calls answered by a given officer removes other MCR officers' calls unevenly across busy and quiet periods, so the call mix becomes correlated with the instrument. However, changes in assigned officers across days within a shift, changes in the shift assignment schedule itself over the program period (Figure A1), and changes over time in who is assigned to staff which day-specific shift, introduce variation

in which officers are scheduled to work during any given time block. The longest time block that includes no scheduled shift changes, under any of the four shift schedules, is 2–8pm. We therefore test for heterogeneity in program impact across unique combinations of officers who were assigned to work during 2–8pm on weekdays (only one weekend combination has an adequate first stage). Because MCR positions were filled by seniority rather than by interest or relevant skill, leaving a mix of more and less motivated officers, similar impacts across sets of scheduled officers would suggest that program practices and norms may be enough to homogenize behavior across a wide variety of officers.

Because heterogeneity analysis requires re-estimating Equation 1 in data subsets rather than with interactions—necessary to properly residualize our instrument in each case—we gauge statistical differences in coefficients across categories using pooled standard errors. That is, for estimates $\hat{\beta}_a$ and $\hat{\beta}_b$ from non-overlapping subsamples a and b, we compute $z = (\hat{\beta}_a - \hat{\beta}_b)/\sqrt{\widehat{se}_a^2 + \widehat{se}_b^2}$ and compare it to standard normal critical values. Because the subsamples share no observations, the two estimates are independent and the variance of their difference is the sum of their variances. To test for heterogeneity across more than two subsamples (the scheduled officer combinations), we use Cochran's Q test of homogeneity, the multi-group analogue of this test. For estimates $\hat{\beta}_k$, $k = 1, \ldots, K$, from non-overlapping subsamples, $Q = \sum_k w_k(\hat{\beta}_k - \bar{\beta})^2$, where $w_k = 1/\widehat{se}_k^2$ and $\bar{\beta} = \sum_k w_k \hat{\beta}_k / \sum_k w_k$ is the precision-weighted mean. Under the null of equal effects, Q is distributed χ^2_{K-1}. With $K = 2$, $Q = z^2$, so the two tests coincide.

3.3 Instrument Validity

This identification strategy recovers the causal, local average treatment effect (LATE) of MCR response on dispatch outcomes if several assumptions hold. First, the instrument must be **relevant**. Figure 5 provides a visual representation of the first stage relationship between our instrument and MCR response. The solid, downward sloping line depicts fitted values from a local linear regression of MCR response on the MCR instrument (number of mental health or suicide related dispatches in 50 minutes prior to dispatch i from other precincts), after partialling out diurnal and annual trends from both variables. Statistical tests also demonstrate relevance (column 1 of Table 2). One additional mental health dispatch in the last 50 minutes outside of own precinct significantly reduces the likelihood of an MCR response by 0.41 percentage points (or 5.5% of the mean MCR response). The Kleibergen-Paap rk LM test rejects underidentification. The robust Kleibergen-Paap Wald F-statistic is also well above the 16.38 critical value for a 10% maximal IV size (Stock and Yogo, 2005), indicating the instruments are not weak.

Second, our instrument must be "as good as" **randomly assigned**. If this is the case, observable call characteristics and dispatcher call-coding choices will not vary with the identifying variation in the instrument. Balance tests support this: although the characteristics of call i are strongly predictive of endogenous MCR response, they are almost universally unrelated to the instrument, conditional on F_{hw}, F_m, t and t^2 (Table 3). Only one call characteristic — the call being from Netcare — is significantly related to the instrument. Jointly these characteristics are conditionally exogenous to the instrument,

with an F-statistic of 1.060.[25] Dispatcher choices regarding priority code (urgency) and police code (topic) are also jointly, conditionally exogenous to our instrument, with F-statistic of 1.56 (Appendix Table A4). In fact, Kerwin *et al.* (2026) demonstrate that these F-statistics are almost surely over-stated, as are all balance test F-statistics.

We similarly test to see if the instrument is exogenous to weather shocks, holidays, lunar cycles, and football games, some of which have been linked to mental health in the literature (Appendix Table A5). Balance test shows that, indeed, these events are correlated with the instrument's deviation from predictable temporal patterns. However, this is not the variation lending first stage predictive power (Appendix Table A6), and therefore not a concern. That is, these time-varying factors are not jointly endogenous to both the instrument and our endogenous treatment variable, MCR response.

Third, for the causal interpretation of the LATE, the instrument must also obey the **exclusion restriction**. That is, in our case, recent mental health dispatch volume outside of one's precinct must have no causal impact on outcomes through any channel other than that of MCR response. This assumption seems viable given the structure of the dispatch process. The MCR decision is made by the dispatcher, who through much of our sample period makes only two other decisions that might influence outcomes – call priority and police code. These codes are conditionally exogenous to the instrument (Appendix Table A4).[26] All later interactions with the individual in crisis are handled by first responders. A responding patrol officer is unaffected by the instrument value by construction.[27] Responding MCR teams may be intuitionally aware of the instrument value, since they do monitor mental health calls city-wide. If their behavior is impacted by the instrument value directly – i.e. though stress levels – this would constitute a violation of the exclusion restriction. Unfortunately there is no excellent way to test this.

In June 2021, dispatchers began making a third decision: whether to route mental health calls through the Right Response Program for clinician telephone consultation. Because Right Response capacity was constrained just as MCR capacity was, and because the same dispatcher made both decisions, our instrument might plausibly have provided a strong first stage for Right Response engagement. It does not (Table A7): the instrument has no predictive power for Right Response engagement, neither in the full sample nor within the lower-severity calls for which the program is primarily used. The Right Response therefore poses no threat to identification—though it is important context, as we'll discuss below.

Last, under heterogeneous treatment effects, a causal interpretation of the 2SLS estimand requires **monotonicity**. In our setting, strict monotonicity requires that any call receiving an MCR response under recent mental health call volume v would also have received one under a lower volume $v' < v$, and that any call not receiving an MCR response at v would not receive one at $v'' > v$. Defiers — calls made *more* likely to receive MCR response by an influx of calls elsewhere in the city — are difficult to rationalize

[25] We created this set of indicators to capture primary dimensions of variation across these calls. There are certainly many other potential indicators not included in these balance tests.

[26] One might argue that how they record the dispatch notes might also be influenced by the instrument, but characteristics pulled from those notes are also conditionally exogenous to the instrument (Table 3).

[27] The precinct leave-out construction of the instrument means it is unrelated to conditions on the ground for the precinct officer handling call i.

when citywide MCR capacity is fixed at one to three teams.[28] Frandsen *et al.* (2023) also show that "average monotonicity" suffices for 2SLS to recover a proper weighted average of individual treatment effects: each call's potential MCR response need only co-vary negatively with the instrument across its values, even if strict monotonicity fails for particular pairs of instrument values. A testable implication of both conditions is that the first stage should be negative, or zero, for any sub-sample defined by a pre-determined characteristic. This is indeed the case if we split the sample by call or dispatch characteristics (Table A8) or by call type, call severity, program eras, or neighborhood category (Table A9); no subsample coefficient is significantly positive. This weighted-average interpretation also requires that our temporal controls correctly capture how the mean of the instrument varies over time (Blandhol *et al.*, 2026). After partialling out our controls, weekend-by-hour, month, and year fixed effects explain only 0.5% of the remaining variation in the instrument. Furthermore, we later show that adding these fixed effects to Equation 1 leaves our core estimates essentially unchanged.

Under either strict or average monotonicity (Frandsen *et al.*, 2023), and provided our temporal controls capture how the instrument varies over time (Blandhol *et al.*, 2026), our 2SLS estimates are a weighted average of treatment effects among compliers; non-compliers receive no weight. Three features of the program shape who these compliers are. First, MCR teams operate citywide only between 10 a.m. and midnight; overtime shifts and late follow-ups on overnight calls are rare, and in these cases MCR availability is decoupled from recent mental health call volume. Consistent with this, the first stage is zero outside these hours (Table A10, Columns 1-2). Second, dispatchers may hold low-priority calls until an MCR team becomes available, decoupling MCR dispatch from recent call volume; consistent with this, the first stage is strong for medium- and high-priority calls but weak and statistically insignificant for low-priority calls (Table A10, Columns 3-4). So, overnight calls and most low-priority calls are never-takers. Third, the program is severely capacity constrained: given the limited number of MCR teams, no group of callers or category of call can always receive an MCR response, so always-takers are unlikely or extremely rare, and, as discussed above, defiers are improbable.[29] Because almost no call can receive an MCR response when teams are unavailable, a call's sensitivity to MCR availability largely reflects how strongly dispatchers and MCR teams prioritize it when teams are available. Within the bounds set by the first two features — medium- or high-priority mental illness incidents between 10 a.m. and midnight — compliers are thus the calls to which MCR teams are triaged, rather than the "marginal" calls they would be in a less capacity-constrained program.

[28] The most plausible mechanism, dispatchers turning to MCR when patrol officers are stretched thin, is blocked by the precinct leave-out construction of the instrument.

[29] MCR response may be more likely at urgent incidents that trigger many calls, but the CAD system merges those calls into one record, so they should not show up as multiple dispatches in our data.

4 Individual-Level Program Impacts

4.1 Average Impacts

Instrumental variables estimates suggest that MCR response has no significant effect on arrival time, but that it roughly quadruples the responder time spent on scene (Columns 1 and 2 of Table 4). Applied to the geometric mean time at dispatch across all patrol-handled calls (32 minutes), this implies roughly 98 additional minutes. However, because the calls MCR teams are triaged toward may differ from the average patrol call, a more relevant baseline is the geometric mean for compliers under patrol response (18 minutes), estimated following Frandsen *et al.* (2023). This implies that MCR response increases the time responders spend on scene by roughly 54 minutes. Even this estimate is imprecise: holding the complier mean fixed, the 95% confidence interval ranges from roughly 9 to 176 additional minutes, and the complier mean is itself an estimate subject to sampling error. Whichever baseline is used, the causal impact of MCR response on time spent on scene is larger than a simple OLS regression would suggest, as shown by comparing Panels A and B of Table 4.

On average, MCR response has no significant effect on the likelihood that an individual in crisis is either left at home (i.e., no transport to jail or hospital, Column 3, Panel A) or voluntarily transported to a hospital (Column 4, Panel A). MCR dispatch does significantly increase the likelihood of transportation to a private psychiatric hospital by 4 percentage points (traditional police transport patients to these hospitals in just 0.2% of cases), but there is no significant impact on voluntary transportation to the free crisis center, to an emergency department without a psychiatric unit, or to one with a psychiatric unit (Columns 5-8, Panel A). In almost every case, OLS coefficients would dramatically misrepresent causal impacts (Panel B).

On average, MCR response has no significant effect on the likelihood that an individual in crisis is either left at home (i.e., no transport to jail or hospital, Column 3, Panel A) or voluntarily transported to a hospital (Column 4, Panel A). These null effects apply to calls in which healthcare transport is already common under patrol response: compliers' estimated likelihood of voluntary transport under patrol response is 35%. MCR dispatch does significantly increase the likelihood of transportation to a private psychiatric hospital by 4 percentage points (from a base of essentially zero transport under patrol response, as shown by both the patrol officer mean and the complier mean), but there is no significant impact on voluntary transportation to the free crisis center, to an emergency department without a psychiatric unit, or to one with a psychiatric unit (Columns 5-8, Panel A). In almost every case, OLS coefficients would dramatically misrepresent causal impacts (Panel B).

The effects of MCR response on downside risks are presented in Panel A of Table 5. MCR response increases the likelihood of involuntary transport to a hospital by 21 percentage points (Column 1), from a base of 5% across all patrol-handled calls and essentially zero among compliers under patrol response. The increase, in line with the statistically insignificant decrease in the likelihood of being left at home (Table 4), is driven primarily by involuntary transports to emergency departments with psychiatric facilities (Column 4

of Table A11). While these emergency departments do have the infrastructure to handle a person in crisis, it is impossible to characterize this result normatively; involuntary transport and hospitalization is necessary in some contexts, and can be harmful in others (Emanuel *et al.*, 2025; Jones *et al.*, 2022). We later examine how these results vary by call type and severity to better understand the context for these transports.

We find no statistically significant effect of MCR response on the probability of arrest or jailing (Columns 2-3, Panel A of Table 5). We also find no support for the hypothesis that MCR teams differ in their likelihood of filing public-order, drug-related, or addiction-supporting charges—those most commonly seen in behavioral health crises (Table A12). Neither do we find that MCR response influences the likelihood of more serious charges, such as crimes against persons, property, or society (Table A13).

Similarly, we find no significant effect of MCR response on police use of force, as recorded by the police Internal Affairs records (Column 4, Panel A of Table 5). Since these records omit low levels of force, we also examine the impact of MCR dispatch using digitized, paper records of police force (Table 6). These paper records, in theory, constitute the universe of force incidents in 2023 and 2024, capturing both low-level force (Levels 0–1, e.g., handcuffing, taser sparking, empty-hand control, pressure points) and high-level force (Levels 2–8). Oddly, MCR response appears to drive an insignificant 9 percentage point increase in low-level force in 2023 (Column 2), but a significant 11 percentage point decrease in low-level force in 2024 (Column 5). These effect sizes are massive; traditional police use low-level force in less than one percent of mental health dispatches, and complier means under patrol response swing implausibly across years (-5% in 2023, 7% in 2024), suggesting these single-year estimates are dominated by noise. Thus, while lack of data prevents confidence about the impact of MCR dispatch on low-level force, it seems likely that MCR response has no significant effect on higher-level use of force (Column 4 of Table 5, Columns 3 and 6 of Table 6).

Our core estimates are robust to adding weekend-by-hour, month, and year fixed effects to the smooth temporal controls in Equation 1 (Table A14 and Table A15). These fixed effects absorb any temporal variation in the instrument that our spline and Fourier controls fail to capture, and so guard against the 2SLS estimand departing from a weighted average of LATEs (Blandhol *et al.*, 2026). Coefficients remain similar in magnitude and significance across all outcomes, with one exception: the effect of MCR response on wait time rises modestly and becomes significant at the 5% level. This is consistent with arrival wait times having particularly strong hour-of-day structure; the two estimates differ by well under one standard error.

A second concern is that recent citywide mental health call volume might affect outcomes through channels other than MCR response — for instance, through weather or events that both generate calls and shape crisis responses, or through strain on hospitals and responders. Three checks suggest this is not the case. First, our core estimates remain qualitatively identical if we control for temporal variation in weather, moonlight conditions, holidays, or football events (Table A16 Table A17). Second, placebo tests in our two "zero first stage" groups shows no alternate channel of impact between the instrument and any outcome (Van Kippersluis and Rietveld, 2018; Angrist and Krueger, 1994); in almost every case, coefficients are fairly precise zeros (Table A18, Table A19, Table A20, Table A21).

Third, the joint test of strict exclusion and pairwise monotonicity proposed by Frandsen *et al.* (2023) does not reject for six of our ten binary outcomes; it rejects only for three of the voluntary transport destinations and use of force (Table A22).[30] This contrasts with several recent applications of the test, which reject for their main outcomes (Huttunen *et al.*, 2026; Norris *et al.*, 2021; Chyn *et al.*, 2025). Test rejection is consistent with violations of either strict exclusion or strict monotonicity. For these four outcomes we therefore rely on the weaker conditions of average monotonicity and average exclusion, under which 2SLS still identifies a proper weighted average of complier effects (Frandsen *et al.*, 2023), and with which our other tests are consistent. We nonetheless omit the specific transport destinations from the heterogeneity analysis, more for parsimony than because of these test results: the distinction among no transport, voluntary transport, and involuntary transport is more central to our questions than where voluntary transports are taken, and generalizes more readily across city-specific institutional contexts.

In sum, MCR response has two main individual-level impacts, on average: it roughly triples the time spent at dispatch, and it raises the likelihood of involuntary transport by 21 percentage points. It also modestly increases voluntary transport to private psychiatric hospitals. While we consider time spent at dispatch an indicator of quality of care, involuntary transport may represent a harmful channel of program impact.

4.2 Program Scope

Table 7 splits results for our primary set of outcomes by call type, and shows that the increase in involuntary transports observed in Table 5 comes mainly from mental health calls (Panel A) rather than suicide-related calls (Panel B). Mental health calls include those related to depression, psychosis, mania, etc., while suicide-related calls involve suicidal ideation or suicide attempts. For both call types, MCR response raises time spent on the call by a similar amount and (insignificantly) lowers the likelihood of being left at home by about 21–22 percentage points. However, where these individuals go instead differs by call type. For suicide-related calls, the drop in being left at home is offset almost one-for-one by a rise in voluntary transport (though neither effect is significant), and the effect on involuntary transport is small and insignificant. For other mental health calls, by contrast, MCR response (insignificantly) reduces voluntary transport while significantly and dramatically increasing involuntary transport by 29 percentage points, against a mean of under 4 percent. MCR response also increases the likelihood of force being used on mental health calls by 4.6 percentage points (significant at the 10% level). Since physical force is often necessary during involuntary transports, it is possible that these two effects reflect the same process.

Table 8 splits results for the same outcomes by call severity (i.e., risk of harm and/or psychiatric acuity), and shows that the MCR-driven increases in involuntary transport and use of force are concentrated among medium-severity calls (Panel B, Columns 5 and 7). For these medium-severity calls, MCR response also (insignificantly) reduces voluntary transport and increases the likelihood of being left at home, whereas for low- and high-severity calls the reverse is true. Splitting the sample by both call severity

[30] The slope component of the test never rejects, so the treatment effects implied between instrument values are always within feasible bounds; the rejections are driven entirely by the fit component.

and call type reduces first-stage power, but shows that this pattern is driven entirely by mental health calls (Table A23 and Table A24). For mental health calls of medium severity, MCR response reduces voluntary transport by 78 percentage points, increases involuntary transport by 48 percentage points, and increases use of force by 7 percentage points. Splitting mental health calls by risk of harm alone or psychiatric acuity alone, rather than by overall severity, yields a broadly similar pattern; the middle tier of each dimension generally shows the largest effects on voluntary and involuntary transport (Table A25 and Table A26), though first-stage power is weak for several subsamples.

A possible, partial explanation for both sets of results is that MCR teams react differently than standard patrol officers under uncertainty. For low- and high-severity calls involving either clear danger or a very obvious case of psychiatric breakdown, both MCR and traditional police may feel that the "correct" course of action is clear. Similarly, suicide attempts or threats are possibly more straightforward to understand and identify than are other mental health complaints, and the ORC code explicitly lists suicidal intent as valid grounds for emergency hospitalization, giving officers a clear protocol absent for mental health calls. Thus, ambiguity might explain why the largest program impacts are seen for non-suicide mental health calls and medium-severity calls. To probe this hypothesis, we examine program impact separately for calls that were coded correctly by dispatchers and for those that were not,[31] with the supposition that incorrectly coded calls are more likely to be ambiguous situations. Indeed, MCR response has an almost identical impact within these incorrectly coded calls as it does within medium-severity mental health calls (Table A27), though 24% of incorrectly coded calls are suicide related and only 38% are medium-severity. The ambiguity explanation does not clarify, however, why MCR response seems to favor *involuntary* transport in situations with ambiguity, rather than voluntary transport.

4.3 Programmatic Era

Table 9, Panels A-C split results by programmatic era (2018-2020, 2021 as a transition year, and 2022-2024). If anything, the effect of MCR response on involuntary transport rises in the later era — a result initially, seemingly at odds with the hypothesis that Netcare clinicians pushed up involuntary transports inappropriately, and city-employed clinicians reduced this propensity. However, the two eras were characterized by a second programmatic change: a reduction in low-severity mental illness calls being dispatched in the later era (Figure 1), due to the advent of the Right Response program in 2021. Involuntary transports rose from 5.0% to 5.8% of police calls across these two eras, in line with the fact that the severity of the mental illness incidents they were handling rose (Figure 2). While MCR teams actually reduced their rate of involuntary transports in 2022-2024, relative to a peak in 2019/2020 (Figure A6), it is possible that the *impact* of MCR response on involuntary transports simultaneously rose if a larger share of the calls driving our estimates — the compliers, to whom the MCR teams are triaged — were medium-severity mental health calls in the later era.

[31] That is, "correctly" coded calls regarding suicide (based on dispatcher remarks) were coded as suicide related, and "correctly" coded calls regarding mental health (based on dispatcher remarks) were coded as mental health related.

Table 10 reports the composition of compliers by call type and severity in each era, weighting calls by their implicit weight in the 2SLS estimate (Frandsen *et al.*, 2023), and shows that compliers did not shift toward medium-severity mental health calls. The complier share of medium-severity mental health calls instead falls from 25% in 2018-2020 to 18% in 2022-2024, a sharper decline than in their share of all calls (26% to 25%). In 2018-2020, complier composition closely mirrors the composition of all calls. In 2022-2024, it does not: high-severity suicide calls rise from 13% to 31% of compliers, despite making up only 14% of calls. That is, in the later era, MCR response was disproportionately triaged towards high-severity suicide calls. This change in triaging was not driven by the Right Response program (Table A28). See Appendix C for details on our adaptation of the Frandsen *et al.* (2023) method.

To test whether this shift in complier composition explains the larger involuntary transport impacts in 2022-2024, we reweight the 2022-2024 sample so that the 2SLS estimate draws on call types and severities in the same proportions as in 2018-2020.[32] We group calls into five cells: low-severity calls (mental illness and suicide pooled) and medium- and high-severity calls separately by call type. Each 2022-2024 call in cell c receives weight $s_{c,1}/s_{c,3}$, where $s_{c,1}$ and $s_{c,3}$ are the complier shares of cell c in 2018-2020 and 2022-2024 (Table 10; the pooled low-severity share is the sum of the two low-severity shares, since the low-severity suicide share is too small to be stable).

Reweighting does not close the gap between eras (Panel D of Table 9): the reweighted 2022-2024 effect on involuntary transport is 0.42, compared with 0.39 unweighted and 0.25 in 2018-2020. So, if anything, holding complier composition by call type and severity constant widens the difference in involuntary transport impact between eras, though the difference is not statistically significant in either the original or the reweighted comparison. Reweighting does, however, close the (also statistically insignificant) gap between eras in the effect on voluntary transport.

Since complier composition does not explain the gap, the effect of MCR response on involuntary transport must itself have changed within call type and severity; Figure 6 suggests that this is the case. Splitting the sample by era and call type-severity cell sharply reduces precision, and several cells have weak first stages (F-statistics and point estimates in Table A29), but the estimates indicate that in 2022-2024 the effect rose for low-severity calls and for high-severity mental health calls, and fell to roughly zero for high-severity suicide-related calls. Combining these cell-level effects with the complier shares in Table 10, the increase in overall impact can be roughly attributed to two cells: low-severity calls, whose effect rose sharply even as their complier share fell, and high-severity mental health calls, whose effect and complier share both rose.[33] Together, these two changes would have roughly doubled the effect of MCR response on involuntary transport between eras, had they not been partly offset by smaller changes in effects for medium-severity mental health calls and high-severity suicide-related calls.

[32] For a related reweighting by complier shares, see Bhuller *et al.* (2020).

[33] A cell's contribution to the change in overall impact can be approximated as its 2022-2024 complier share times its 2022-2024 effect, minus the same product for 2018-2020; for low-severity calls, $(0.18 \times 0.59) - (0.30 \times 0.03) \approx 0.10$. Because each cell's effect is estimated within its own subsample (which weights compliers differently than the pooled estimate), often with a weak first stage, these contributions do not sum exactly to the difference between Panels A and C of Table 9.

Jointly, these results suggest that program impacts depend on details of implementation and the surrounding policy environment. The Right Response program, by diverting many low-severity mental illness calls from dispatch after 2021, changed the calls MCR teams handled, and changes in triage shifted MCR toward high-severity calls even beyond this change in call mix. The switch from third-party clinicians to city-employed clinicians, who have no legal authority over involuntary transport but may still influence it, likely shifted program norms. The local healthcare landscape mattered too: although year-by-year estimates are imprecise, MCR response shifted voluntary transports toward private psychiatric hospitals and away from emergency departments with psychiatric units in 2022 and 2023 only (Figure A8), years during which the closure of the free Crisis Center and another large, public hospital in the south of Columbus (Mount Carmel West) had left public hospitals overwhelmed with psychiatric cases. In this period, the ADAMH Board of Franklin County, which coordinates behavioral healthcare, had asked MCR leadership to transport more patients to private psychiatric hospitals, a few of which had in turn agreed to accept more uninsured and Medicaid patients. The effects of co-responder programs are therefore not fixed properties of the general model itself, but depend on who staffs them, which calls reach them, and what services exist to receive their patients.

4.4 Demographic and Spatial Inequality

Alternative crisis response is often motivated by a desire to "fix" underlying inequalities in the quality of default crisis care; Figure 7 suggests that such inequalities in default police response do exist in Columbus (neighborhood categories are mapped in Figure A2). On average, police spend 5–6 more minutes on scene for crises in higher-income neighborhoods than for crises in low-income neighborhoods, and these crises are less likely to end in arrest or jailing. On the other hand, crises in higher-income neighborhoods receive the slowest police response and are the most likely to end in involuntary transport, while crises in low-income, non-White neighborhoods are the most likely to end in voluntary transport — patterns that fit less cleanly with a narrative in which higher-income households receive better healthcare and less coercive treatment. These means and differences remain essentially identical when adjusted for call type and crisis severity. MCR teams themselves also respond more often to crises in higher-income neighborhoods, accounting for 8.2% of mental illness dispatches in our sample, versus 6.9% and 7.2% in low-income, non-White and low-income, White neighborhoods (bottom-right panel of Figure 7).

Table 11 splits MCR impacts by neighborhood type, and shows that MCR response increases voluntary transport by 34 percentage points in higher-income neighborhoods, while having no significant effect (and negative point estimates) on voluntary transport in low-income neighborhoods. This spatial difference is statistically significant. In wealthier neighborhoods, this increase in voluntary transport is paired with an (insignificant) increase in involuntary transport, and together the two drive a dramatic, 57 percentage point reduction in the likelihood of being left at home; that is, MCR response *adds* medical transports. The impact of MCR response on involuntary transport (insignificant once split by neighborhood) is similar in low-income neighborhoods, but there it is paired with an (insignificant) reduction in voluntary transport and no reduction in the likelihood of being left at home; that is, in low-income neighborhoods, MCR response *substitutes* involuntary for voluntary transports. This substitution is starkest

in low-income, non-White neighborhoods, with effects roughly half as large (though not statistically indistinguishable) in low-income, White neighborhoods.

In addition, MCR response dramatically increases time spent with those in crisis in higher-income neighborhoods, while having no statistically significant effect on time spent in lower-income neighborhoods (Column 2 of Table 11). The additional time spent in higher-income neighborhoods (a difference that is not statistically significant) might reflect a mechanical increase in the time spent on transportation; we cannot separate time spent on scene from time spent on transport. It might also reflect the de-escalation effort necessary to secure voluntary transport, or other dimensions of quality of care. It could also reflects cultural norms and power dynamics; wealthier families may feel empowered to demand a greater quantity of first responder time than do lower-income families.

Table 12 further splits MCR impacts on transport location by neighborhood type, pooling voluntary and involuntary transports. When answered by patrol police, crises in higher-income neighborhoods are less likely than crises in low-income neighborhoods to end in transport to the free crisis center, but more likely to end in transport to an emergency department with a psychiatric unit (Figure A9). In higher-income neighborhoods, MCR response increases the likelihood of both types of transport (Columns 1 and 4 of Table 12). This could reflect MCR teams' willingness to transport individuals to these facilities even when they are relatively far away, as crises in higher-income neighborhoods tend to be (see mean distances in Table 12); however, patterns by distance are weaker than patterns by neighborhood type (Table A31). Because upper-class families are anecdotally reluctant to go to the free crisis center due to stigma, it is possible that discussion with MCR team members erodes that stigma, thereby equalizing the rate of transport across neighborhood categories. The impacts on transports to emergency departments with psychiatric units is harder to explain, and could be a statistical fluke; the difference across even low-income, non-White and wealthy neighborhoods is not statistically significant at traditional levels.

All in all, we find that MCR response widens some neighborhood disparities, narrows others, and leaves others roughly unchanged. MCR response raises time spent at dispatch most in higher-income neighborhoods, where police already spend the most time, and so likely widens this disparity. It raises voluntary transport only in higher-income neighborhoods, where police transport individuals voluntarily less often than in lower-income non-White neighborhoods, and so likely narrows and may reverse this disparity, which is admittedly perplexing to begin with. The effect of MCR response on involuntary transport is similar across neighborhoods, though in lower-income neighborhoods these involuntary transports partly replace voluntary ones rather than adding to them. MCR response narrows the disparity in transports to the free crisis center, though such transports are rare in all communities. More importantly, given the dominance of emergency department transports, it widens the disparity in transports to emergency departments with psychiatric units by raising such transports primarily in wealthy communities. Finally, because MCR teams respond more often in higher-income neighborhoods, each of these effects reaches a larger share of dispatches there.

4.5 Learning and Person-Specific Skills

Table 13 splits results for dispatches by MCR officers who have worked in the program for less than two years and dispatches by more experienced MCR officers. Effect sizes are statistically consistent across the two panels. While MCR response appears to lead to a marginally significant increase in use of force for experienced officers (Panel B), which is not the case for less experienced officers (Panel A), the two coefficients cannot be distinguished statistically. Calendar time, which is mechanically correlated with program experience, is controlled for as in all specifications (see Equation 3). However, conditional on these controls, experience in the MCR program is correlated with officer tenure; less experienced MCR officers have on average six less years of department tenure. Controlling for department tenure is nonsensical (it is itself an outcome of MCR response), and thus the panels are conceptually split by both measures, along with any other correlates. Splitting results by department tenure rather than MCR experience leads to similar homogeneity, with no statistically significant impact on use of force for more experienced officers (Table A32).

Turning to the question of heterogeneity across individual officers, there are 60 unique, scheduled officer combinations observed within the 2–8pm window on weekdays, and 30 on weekends. Many of these combinations have few observations, however. We drop all combinations with a first-stage F-statistic below 5, leaving only six weekday combinations. Cochran's Q test p-values range from 0.17 to 0.88 across outcomes (Figure 8). While this is consistent with program norms and training outweighing any effect that individual tendencies, beliefs, or skill-sets might have on program impacts, the test has limited power. All but one of the six remaining combinations have a first-stage F-statistic below 10, and confidence intervals are wide enough that meaningful differences across scheduled officer combinations cannot be ruled out.

5 City-Level Program Impacts

While advocates of alternative response programs generally cite a need to improve healthcare for those in crisis and/or reduce police violence – both outcomes at the individual level — they may also argue that improved crisis response will cut city costs in the short, medium or long run by reducing repeat calls via more effective healthcare provision, by reducing instance of jailing, and/or by reducing the publicly born, future healthcare costs associated with untreated mental health conditions. Of course, if alternative response programs improve trust in the 911 system, they might also increase call rates and expenditures on crisis response — which might in turn increase medium-run healthcare costs through additional hospitalizations and/or reduce long-run healthcare costs via improved preventative care. Occasionally, advocates make another claim: that taking mental health response off the hands of patrol police allows them to respond more swiftly and concentrate more effectively on other types of calls, improving the daily experience of police officers as well as the services they provide to the public.

We next investigate these city-level program impacts using the same identification strategy as before (Equations 1 and 3), and outcomes that are sometimes defined at

the dispatch or call level (Y_{ipt} as before), and sometimes at the police precinct level (Y_{pt}).

5.1 Resource Utilization

First, we examine the immediate salary cost of MCR response. The direct cost of the MCR program depends on whether it changes the number of units (cars) or responders (officers and/or social workers) involved in an incident, changes the salaries of those responders, and/or changes response periods. Salaries are determined by rank; sergeants make 18% more per hour than the top of the hourly range for officers, and lieutenants make 18% more than sergeants. We use unit-specific dispatch and closing times to calculate total responder-hours spent on scene (including the time spent by officers and by MCR clinicians) for each call in our dataset.[34] We also assign salary ranges to each responder, based on their rank, as indicated by their unit number. Salary data for police and for clinicians was shared by the Columbus Police Department for 2024.

MCR response increases both the number of units who respond to a call and the number of on-scene responders (Table 14, Columns 1–2). Responding units increase by about one and a half units (from a base of 2.2 across all patrol-handled calls and 1.6 among compliers under patrol response), and responders by 1.9 individuals (from a base of 2.6 under either heuristic) – suggesting that the two-person MCR teams are generally adding to patrol response rather than substituting for patrol officers. This makes some sense, since in 82% of the MCR dispatches in our sample data, MCR co-respond with police. MCR response also increases total responder-hours, roughly fivefold (Table 14, Column 3), unsurprising since it increases the number of responders and the total length of time spent by any responder at dispatch (Table 4). This rules out a complete substitution of MCR time spent on scene with patrol officer time spent on scene, however. In fact, there is no significant effect of MCR response on patrol officer-hours (Column 4). Together, the increase in responders and in responder-hours roughly quintuples the salary cost of response (Table 14, Column 5), though this estimate is imprecise.

MCR teams generally respond to calls whose police codes require a minimum of two officers according to department policy (99% of calls MCR responds to fall in this category, and 95% of calls in our sample do), which could be part of why we don't observe substitution of MCR responders for patrol responders – the clinicians do not count towards this requirement. If a second, dispatched unit includes two officers when only one more is required, this would push up the number of responding police officers even beyond requirement. Table A33 finds no significant impact of MCR response on the total number of officers dispatched, though the point estimate suggests that on average 0.7 more officers are dispatched as a result of MCR response (Column 1), none of whom

[34] This requires imputing the number of officers in each unit; we know the number of officers and the number of units that responded to each event, but not which units include which officers. Sergeants and lieutenants essentially always respond alone, and their units are distinguishable in the data; MCR units, also distinguishable, never include more than 1 officer. We assign 1 officer each to those units as well as to a few other special purpose units that almost always contain 1 officer. Remaining officers on the call are assumed to be responding within the patrol units, or a few other types of multi-officer response units we group with patrol units. Because all of these patrol and multi-officer units are assigned the same salary, the true distribution of officers across this units does not matter for salary cost estimates.

are patrol officers (Column 2) and 0.4 more officers who are unrequired by the policy (Column 3). So, the MCR clinician accounts for one of the 1.9 additional responders, and the MCR officer for the remainder, with no significant substitution with or increases in patrol officers. The point estimates in columns 1 and 3 of Table A33 differ because for 10% of sample incidents, fewer officers are dispatched than required; it seems that MCR response pushes the average incident toward meeting this requirement.

These estimates can be translated into extra salary dollars spent, per dispatch and per year. For compliers, the geometric mean salary cost of patrol response is roughly $35 per call (Table 14, Column 5); MCR response multiplies this by an estimated 4.9, implying an additional salary cost of roughly $135 per call (95% confidence interval: $26 to $444, holding the complier mean fixed). With MCR teams responding to an average of about 1,700 calls per year in our sample, this amounts to roughly $234,000 in additional responder salary costs per year (95% confidence interval: $44,000 to $769,000). Several caveats apply. These figures describe compliers in our sample, the calls to which MCR teams are primarily triaged, and are not average treatment effects across all mental illness calls. They also cover only MCR responses to 911 and non-emergency calls; as discussed above, we drop MCR dispatches to calls the teams received on their work phones. The time use and salary implications of these dispatches are likely quite different, because MCR teams respond alone (without patrol officers) to these direct calls 68 percent of the time, as opposed to 18 percent of the time in our sample. However, such calls make up only 3.5 percent of MCR dispatches, so excluding them is unlikely to change the overall cost estimate substantially. Finally, these estimates value only the responder time spent on these calls, and exclude other costs such as fuel or vehicle wear.

5.2 Repeat Calls and Hospitalizations

Next, we examine the likelihood of repeat calls from caller i, defined as repeat calls from the same civilian phone number. (Institutional callers are dropped for this analysis.) Repeat calls are common in our sample data; 18 percent of callers will call again about a behavioral health (i.e., mental illness or substance use) related crisis within the next 30 days, and 18 percent will call again about a non-behavioral health related crisis. Six percent will call again for both reasons. Of those who call again within a month, 57 percent call more than once, and 19 percent call more than 10 times. Almost half — 48 percent — of callers will call again within 180 days, calling an average 4.4 times in that period. About half of those calls relate to mental illness. Given that the average response in our larger dispatch dataset costs $114 in salary, and the average mental illness response in our sample data costs $143, it is perhaps unsurprising that Columbus city officials hoped that the MCR program would reduce repeat call rates (Namigadde, 2019).

We view repeat calls in the short-run (specified as 30 days) as the product of two underlying factors — effective healthcare provision at dispatch i should make a second or third call for that same crisis less necessary, while the experience of a safe, effective, and/or compassionate response may build trust in 911 and increase the likelihood of calling 911 for that same crisis or another crisis, if the need for help does arise again. We expect the dominant factor to be effective healthcare provision in the short run, but both factors exist and are working against one another in their effect. Repeat calls in the

medium-run (specified as 180 days) are the product of the same two underlying factors; while the crisis underlying call i is likely resolved one month later, effective healthcare provision may help to prevent the next crisis, while increased trust may increase the likelihood of calling 911, conditional on another crisis occurring. Because even highly effective crisis response for one particular call is unlikely to prevent all future crises, we expect that in the medium-run, the dominant factor impacting likelihood of repeat calls is trust.

If our hypotheses are correct, and if MCR response both improves healthcare provision and builds trust in 911 crisis response, we should see that it has a negative effect on repeat calls in the short run and a positive effect over a longer horizon. Columns 1 and 2 of Table 15 show that this is the case for civilian callers, though neither effect is statistically significant. The LATE estimate for any repeat call is -5.5 percentage points at 30 days (from a repeat call rate of 18% across patrol-handled calls) and 28 percentage points within 180 days (from 29% across patrol-handled calls).[35] Effects on the number of repeat calls and on the likelihood of a repeat call accompanied by hospitalization within 180 days are also statistically insignificant (Columns 3–4 of Table 15), with a negative point estimate for the number of calls (-0.56, from a complier mean of 2.9) and an estimate of essentially zero for hospitalization. Thus, if MCR increases the likelihood of at least one repeat call, it does not appear to increase the number of repeat calls or the likelihood of hospitalization.

To examine the trust-building mechanism alone, we conduct two additional tests. First, we examine how MCR response affects repeat calls from caller i that do not regard behavioral health. If MCR response builds general trust through better call handling, that trust could spill over into repeat 911 engagement for any call type. Since there is no crisis-resolution mechanism at play for non-behavioral health calls, we should observe a positive impact on the likelihood of both short and medium-run repeat calls, as well as an increase in the intensive margin number of calls. Table 16 demonstrates that these hypotheses play out, though again point estimates are not statistically significant.

Second, we examine how MCR response impacts future calls from within the neighborhood of call i. As before, we examine future calls received within 30 and 180 days of dispatch i, within 50m, 100m, and 200m radii of caller i's location, excluding those of caller i themselves. Again, we examine future calls in the location of civilian callers only. Since the neighbors closest to an incident are more likely to have witnessed how police or MCR handled the call, we expect the impact on future calls to be strongest at the 50m radius and to taper off at 100m and 200m. Because the crisis-resolution mechanism again does not apply here, the trust building effect of MCR response should be isolated. And given the results in Table 16, this trust building effect should extend beyond behavioral health calls. Figure 9 presents results for future, geographically-local calls (regarding behavioral health crises or anything else), with point estimates in Appendix Table A34. The pattern observed is consistent with our prediction: at 30 days, all three coefficients are positive but statistically insignificant, declining from 22.93 percentage points at the 50m radius to 5.2 percentage points at the 200m radius. At 180 days, the estimates follow

[35] Patrol officer means may not reflect repeat call rates among compliers, but complier means for these outcomes are so imprecisely estimated that they imply a lower repeat call rate within 180 days than within 30 days. We report them for continuity only.

a similar declining trend as the radius expands, with the 50m estimate of 20.6 percentage points statistically significant at the 5 percent level.

In summary, MCR response has no statistically significant effect on repeat behavioral health calls overall. This null result could reflect no effect, or two channels working in opposite directions: building trust in 911, which raises future calls, and improving crisis resolution, which reduces future health needs. Several pieces of suggestive evidence are consistent with both channels being at play. MCR response has a negative (though insignificant) effect on the short-run likelihood of repeat behavioral health calls, consistent with crisis resolution, while it (insignificantly) increases repeat calls unrelated to behavioral health as well as increasing future calls from the immediate neighborhood (significantly so within 50m over 180 days), consistent with trust building. Finally, MCR response does not increase the number of repeat behavioral health calls or the likelihood of a repeat call accompanied by hospitalization over 180 days, suggesting that any increase in the likelihood of repeat engagement is not driven by worsening crises.

6 Conclusion

Our analysis shows that Columbus's Mobile Crisis Response program, a co-response between police and clinicians, meaningfully alters the trajectory of mental health 911 calls. Co-response roughly quadruples the time responders spend at dispatch (potentially reflecting de-escalation and assessment), and it increases the likelihood of involuntary transport by 21 percentage points, an increase concentrated among medium-severity mental health calls, which also see greater use of force. On average, co-response has no effect on arrest, jailing, or high-level use of force, and it increases voluntary healthcare transport only under certain conditions.

Contextual details matter for impact. Co-response increases voluntary healthcare transport and time spent at dispatch the most in higher-income neighborhoods, which are also about 1 percentage point more likely to receive co-response, raising concerns about equity. The local healthcare landscape matters too: MCR shifted voluntary transports toward private psychiatric hospitals only in 2022 and 2023, when closures had left public hospitals overwhelmed with psychiatric cases. The employment domain of clinicians may also matter – while some co-responder programs hire third party clinicians, others hire them as city employees. Columbus switched models partway through our sample period, and the effect of co-response on involuntary transport, if anything, rose after the switch to city-employed clinicians, though this change cannot be separated from concurrent shifts in triage, call diversion, and post-Covid mental health.

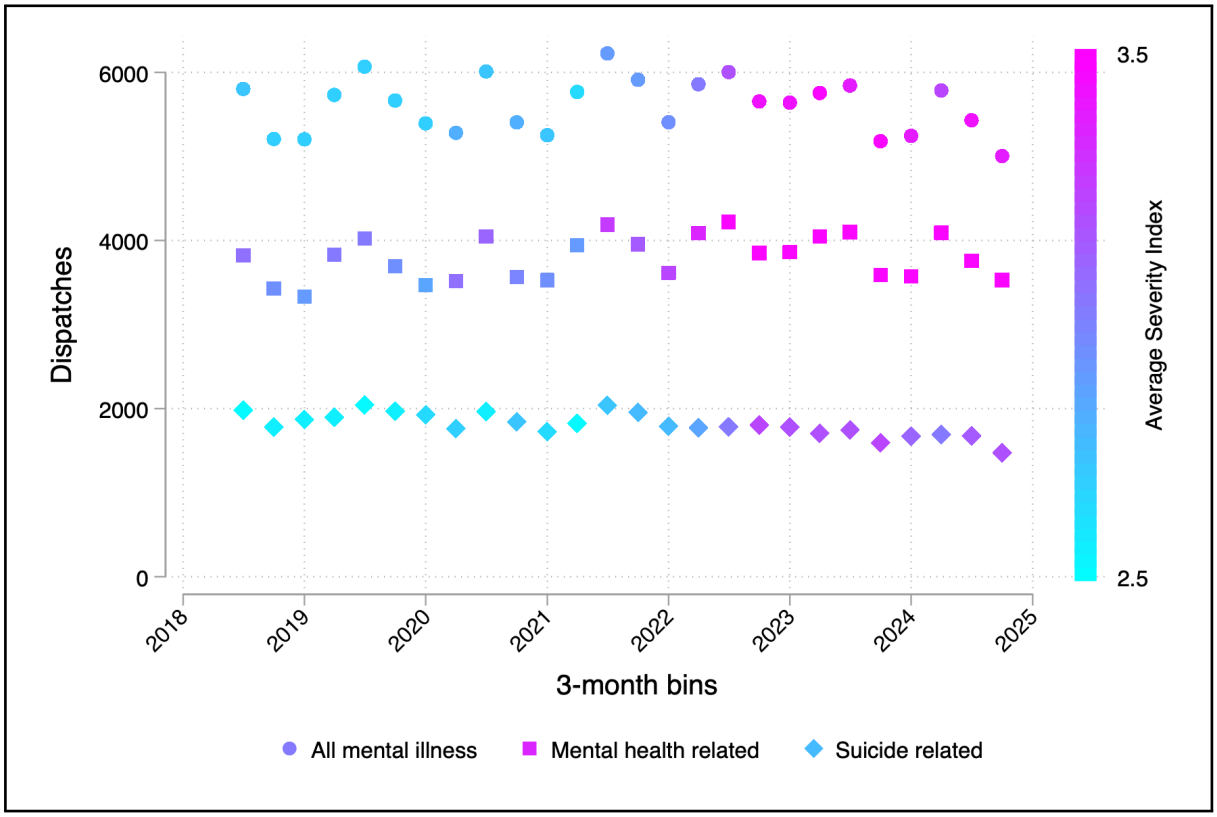
At the city level, the program increases operational costs: co-responders add to rather than replace patrol officers, and responder salary costs per call roughly quintuple. We hypothesized that co-response may build trust in the 911 system, increasing repeat calls, while improved crisis resolution could reduce the necessity for repeat calls. We see no significant effect on repeat calls overall, though suggestive evidence is consistent with both channels being at play.

Our findings also speak to how policymakers should interpret the growing empirical

literature on crisis response. Our findings also speak to how policymakers should interpret the growing empirical literature on crisis response. We provide credibly identified evidence on a stand-alone police-clinician co-responder program—distinct from models in which co-response operates alongside parallel civilian teams (as in Durham's HEART) or in which "non-police" responders frequently co-respond with officers (as with CAHOOTS). That distinction matters: while Davis *et al.* (2025) find that CAHOOTS sharply reduces involuntary mental health holds, co-response in Columbus, where every response includes an officer with the authority to initiate involuntary transport, increases them. Interpreting this shift normatively requires caution. Involuntary transport may represent appropriate clinical intervention for individuals at substantial risk, and MCR's involuntary transports go primarily to emergency departments equipped with psychiatric units. Yet involuntary transport raises long-standing concerns among mental-health advocates, and evidence suggests that for marginal cases it can be extremely harmful (Emanuel *et al.*, 2025). More broadly, our results underscore that the impacts of co-response depend heavily on who is triaged to it, who staffs it, and what services exist to receive its patients. Cities expanding crisis-response capacity should therefore not expect co-response to deliver reliable impacts on its own, but treat it as one component of a broader continuum of care whose effects depend on the rest of that system.

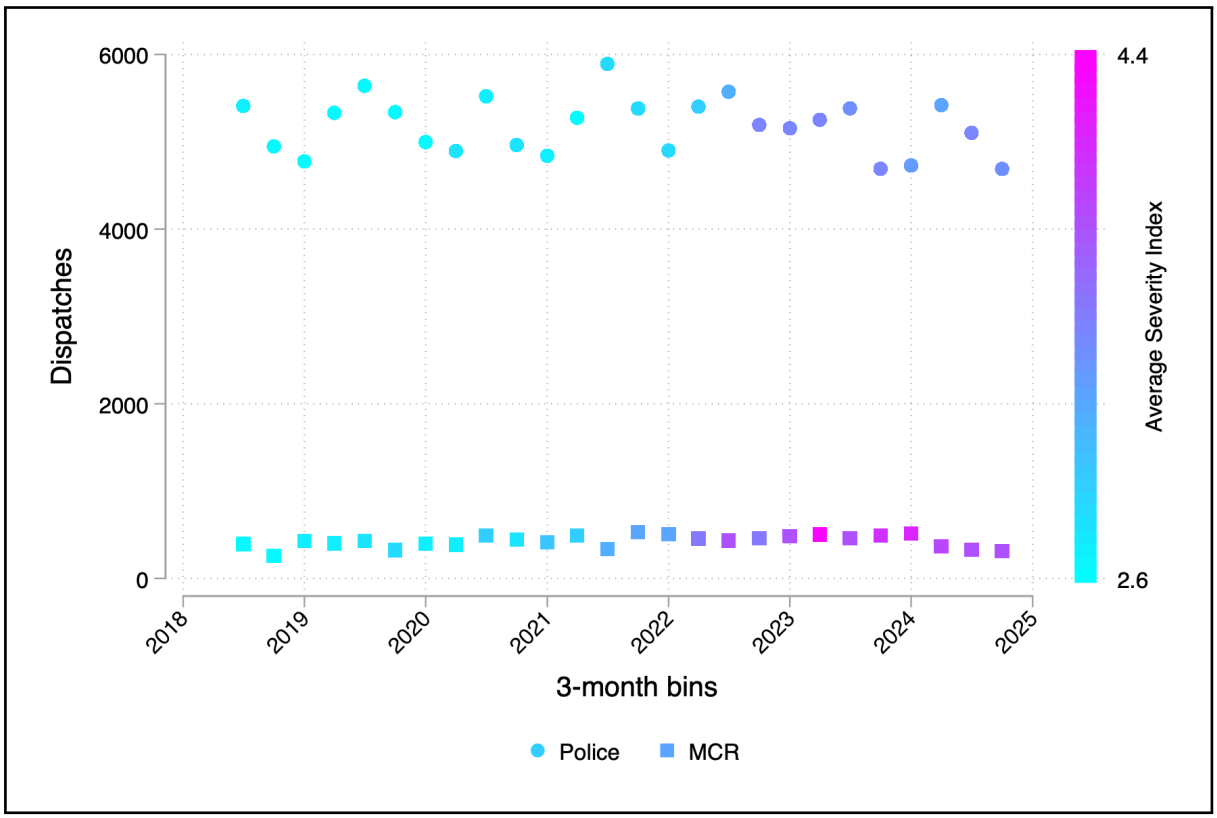
Figures

Figure 1: Trends in Mental Illness Dispatches and Call Severity



Note: The marker color within the legend indicates the average severity. June 2018 is dropped for visual purposes; it is outside the nearest 3-month bin.

Figure 2: Trends in Mental Illness Dispatches and Call Severity by Responder



Note: The marker color within the legend indicates the average severity. June 2018 is dropped for visual purposes; it is outside the nearest 3-month bin.

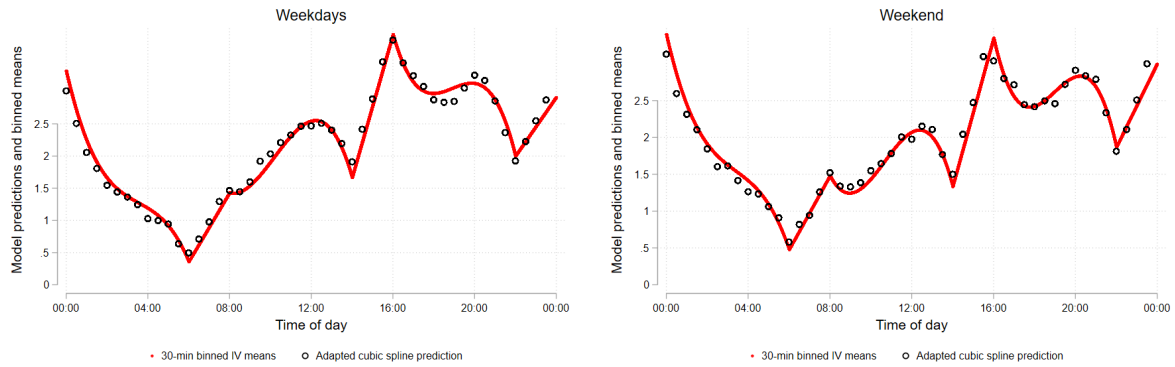


Figure 3: Dispatch Volume Diurnal Trends and Residualization

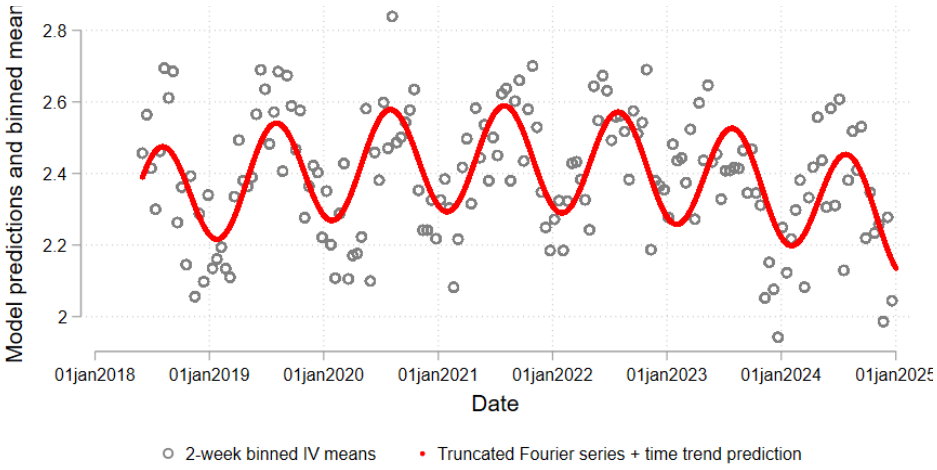
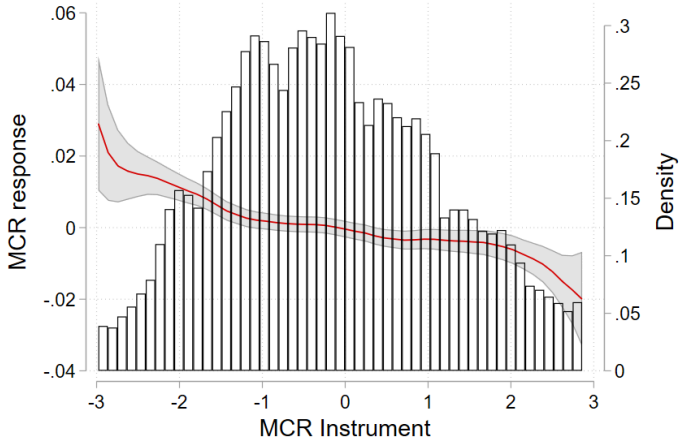


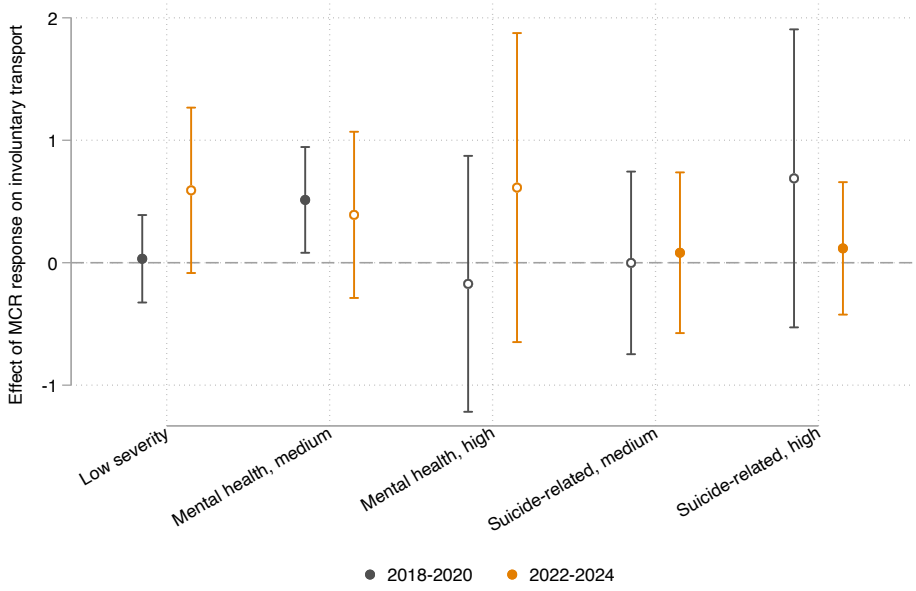
Figure 4: Dispatch Volume Annual Trends and Residualization

Figure 5: MCR First Stage Regression



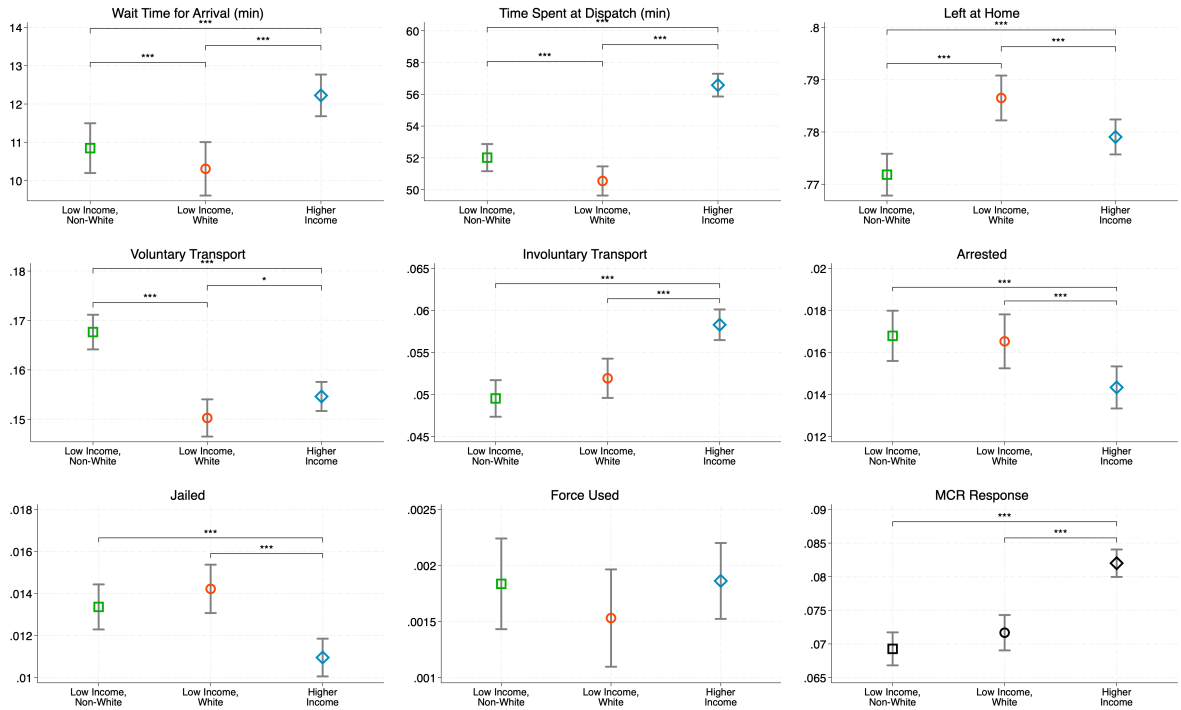
Note: Solid line depicts fitted values from a local linear regression of MCR response on the MCR instrument, after partialling out partialling out diurnal and annual trends from both variables via our primary specification (Equations 1-2). The gray band represents 95% confidence intervals. Histogram displays the distribution of the residualized instrument. For visual clarity, observations below first percentile and above 95th percentile of the residualized instrument were excluded.

Figure 6: Involuntary Transport Impact by Call Type, Call Severity and Era



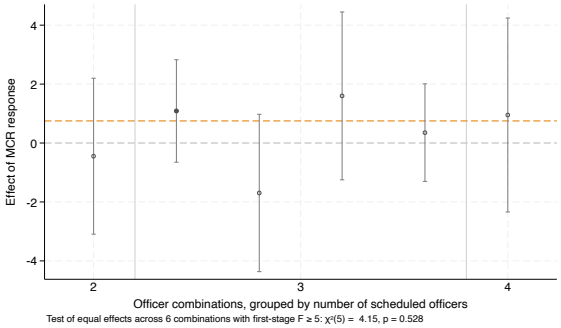
Note: Lines indicate 95% confidence intervals. Hollow markers indicate that the first stage F-statistic (shown in Table A29) is below 10.

Figure 7: Outcome and MCR Response by Neighborhood Category

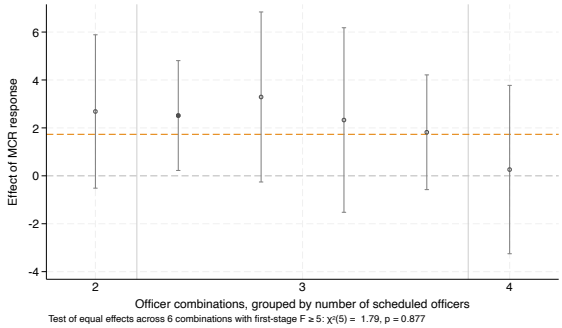


Note: In all subfigures except the bottom-right one, neighborhood-specific means have been calculated in the universe of all mental illness dispatches by standard, patrol police (i.e., excluding dispatches with MCR response). The final subfigure graphs raw MCR response rate by neighborhood for mental illness dispatches. Stars indicate mean difference t-test p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. T-tests of time means were done with log time.

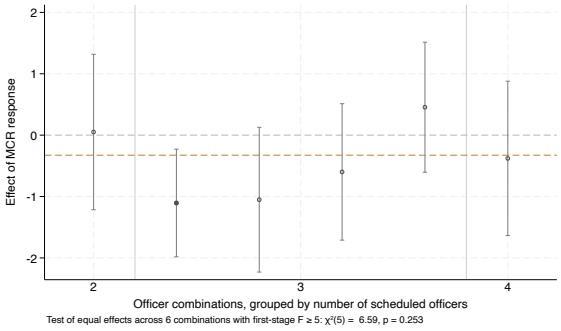
Figure 8: MCR Response Impact by Scheduled Officer Combination, Weekdays 2–8pm



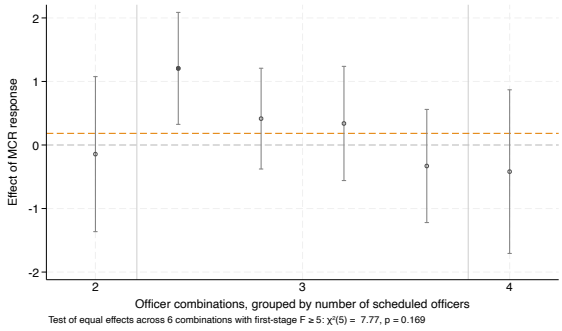
(a) Wait Time for Arrive (log min)



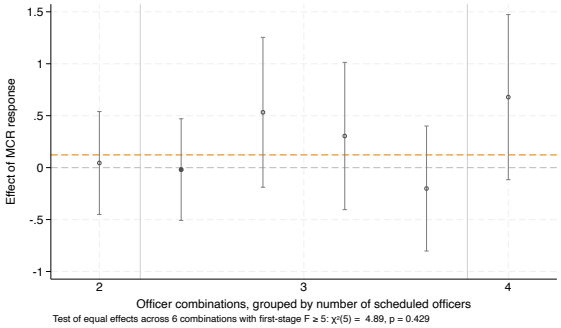
(b) Time Spent at Dispatch (log min)



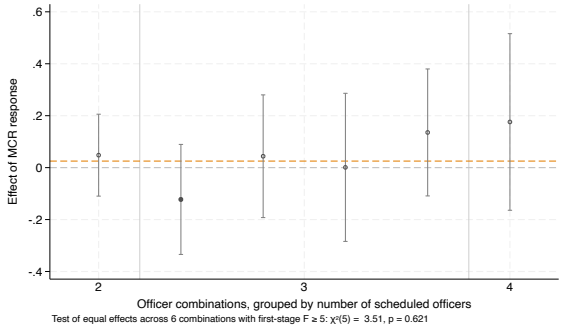
(c) Left at Home



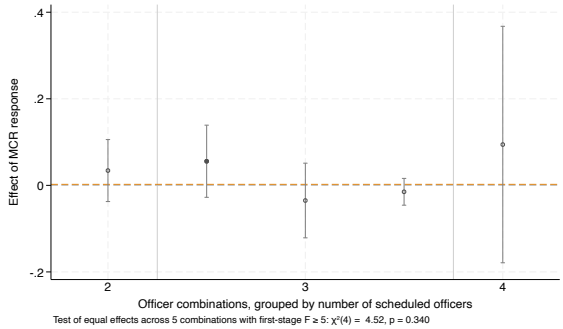
(d) Voluntary Transport



(e) Involuntary Transport



(f) Jailed



(g) Force Used

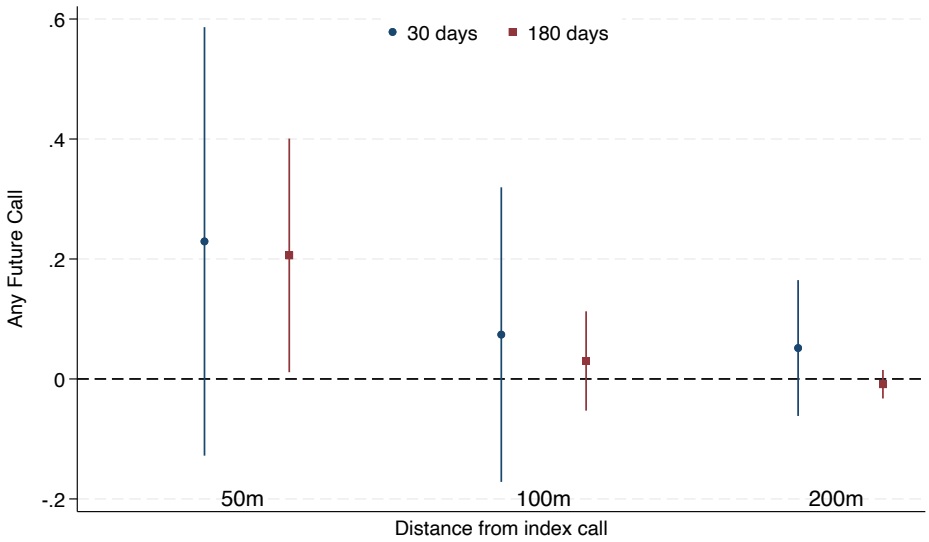


Figure 9: Future Calls from the Neighborhood of Call i

Tables

Table 1: Summary statistics by police response model

	MCR	Traditional Police	Difference
Time to arrive (min)	15.5461	11.3008	4.2454***
	(19.5992)	(67.4726)	(0.2604)
Time at dispatch (min)	119.1270	53.5883	65.5387***
	(92.4871)	(88.7540)	(0.9087)
No transport to jail or hospital	0.5365	0.7788	-0.2423***
	(0.4987)	(0.4151)	(0.0049)
Voluntary Transport to:			
Any hospital	0.2735	0.1574	0.1160***
	(0.4458)	(0.3642)	(0.0043)
Crisis Center	0.0553	0.0124	0.0429***
	(0.2286)	(0.1108)	(0.0022)
Psych Hospital	0.0023	0.0021	0.0003
	(0.0483)	(0.0454)	(0.0005)
Emergency Dept with No Psych Unit	0.0466	0.0575	-0.0109***
	(0.2108)	(0.2328)	(0.0021)
Emergency Dept with Psych Unit	0.1692	0.0854	0.0838***
	(0.3750)	(0.2795)	(0.0036)
Involuntary Transport	0.1816	0.0540	0.1277***
	(0.3856)	(0.2260)	(0.0037)
Arrested	0.0209	0.0157	0.0053***
	(0.1431)	(0.1242)	(0.0014)
Jailed	0.0166	0.0125	0.0041***
	(0.1278)	(0.1113)	(0.0012)
Force Used	0.0045	0.0018	0.0027***
	(0.0669)	(0.0420)	(0.0006)
Repeat calls (30 days)	0.2278	0.2286	-0.0008
	(0.4194)	(0.4199)	(0.0044)
Repeat calls (180 days)	0.3560	0.3521	0.0039
	(0.4788)	(0.4776)	(0.0051)
Repeat Healthcare Contact (30 days)	0.1274	0.1004	0.0270***
	(0.3334)	(0.3005)	(0.0035)
Repeat Healthcare Contact (180 days)	0.1976	0.1602	0.0375***
	(0.3982)	(0.3668)	(0.0042)
Observations	11138	136540	147678

NOTE: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Summary stats in col 1−3 have been computed for mental and suicide related dispatches between June 2018 and December 2024. Mental health crisis centers include facilities like Netcare and RI, private psychiatric hospitals include Twin Valley Behavioral Healthcare Hospital, Columbus Springs (Dublin & East), & Sun Behavioral Health, Torii Behavioral Health (Formerly SEQUEL-Pomegranate Health Systems), RiverVista Behavioral Hospital, and Mount Carmel Behavioral Health. Hospital facilities with an emergency department and psychiatric unit includes Riverside Methodist Hospital, Ohio State University Hospital - Main and Harding, Nationwide Children's Hospital, Big Lots Behavioral Health Pavilion, and Mount Carmel West Hospital (up untill 2019). Hospital facilities with an emergency department and no psychiatric unit includes Dublin Methodist hospital, Grant medical center, Ohio Health systems, Mount Carmel, Doctors hospital, and other centers of Ohio State university hospital.

Table 2: MCR First Stage Regression

	(1) MCR responded	(2) MCR responded	(3) MCR responded
# MH dispatches in 50 min prior to dispatch i from OTHER precincts	-0.0041***		
	(0.0004)		
=1 if > 2 MH dispatches in 50 min prior to dispatch i from OTHER precincts		-0.0114***	
		(0.0015)	
=1 if > 3 MH dispatches in 50 min prior to dispatch i from OTHER precincts			-0.0115***
			(0.0017)
Obs	147,678	147,678	147,678
Weak id KP Wald F	88.3140	57.3308	43.9249
Underid KP LM stat	88.1331	57.3159	43.9076
Underid LM stat p-val	0.0000	0.0000	0.0000
Dep Mean	0.0750	0.0750	0.0750
Weekend × Adapted Cubic Spline	Y	Y	Y
Annual 1st-order Fourier	Y	Y	Y
Time, Time^2	Y	Y	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates are obtained after controlling for diurnal trends, annual trends, and time trends – via our primary specification (Equations 1-2) in Columns 1-3, and via weekend-hour, month, and year fixed effects in Column 4. Controls are identical in Columns 1-3 and 5 except that we use a 6th-order rather than 2nd-order truncated Fourier series for capturing diurnal trends in Column 5. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. We also report the Kleibergen-Paap rk LM heteroskedasticity-robust underidentification test statistic with the corresponding p-value reported below.

Table 3: Instrument Balance on Call Characteristics

	(1) MCR	(2) IV
Call from 911	0.0039***	-0.0110
	(0.0014)	(0.0085)
Call from Netcare	0.0355***	-0.0397**
	(0.0037)	(0.0194)
Call mentioning passerby	0.0201***	0.0353
	(0.0070)	(0.0392)
Call from an anonymous person	-0.0116***	0.0076
	(0.0024)	(0.0159)
Black race mentioned	-0.0078***	0.0035
	(0.0016)	(0.0102)
White race mentioned	-0.0003	0.0087
	(0.0017)	(0.0100)
Hispanic race mentioned	0.0068	-0.0410
	(0.0051)	(0.0301)
Only woman mentioned	-0.0202***	-0.0134
	(0.0039)	(0.0227)
Only man mentioned	-0.0244***	0.0101
	(0.0039)	(0.0225)
Woman and man mentioned	-0.0245***	0.0014
	(0.0040)	(0.0234)
Children mentioned	-0.0085***	0.0149
	(0.0022)	(0.0130)
Family members mentioned	0.0087***	-0.0078
	(0.0016)	(0.0098)
Current/former partner mentioned	-0.0038*	-0.0040
	(0.0021)	(0.0130)
Homeless/panhandling mentioned	-0.0182***	0.0136
	(0.0041)	(0.0277)
Weapon mentioned	0.0188***	-0.0134
	(0.0023)	(0.0130)
Violence mentioned	0.0019	0.0019
	(0.0028)	(0.0165)
Off meds mentioned	0.0146***	-0.0124
	(0.0039)	(0.0207)
Crazy/Insane/mental mentioned	-0.0238***	0.0019
	(0.0032)	(0.0243)
Schizophr./ Bi-polar mentioned	0.0119***	-0.0113
	(0.0024)	(0.0136)
Depression mentioned	-0.0003	0.0245
	(0.0030)	(0.0182)
Obs	147678	147678
Dep Mean	0.0750	2.4010
Rsq	0.0548	0.1731
Joint test F-stat	22.5150	1.0600
Joint test p-val	0.0000	0.3860
Weekend × Adapted Cubic Spline	Y	Y
Annual 1st-order Fourier	Y	Y
Time, Time^2	Y	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates are obtained after controlling for diurnal trends, annual trends, and time trends – via our primary specification in Columns 1 and 2.

Table 4: Time Spent and Healthcare Access (OLS and IV)

					Voluntary Transport to:			
	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport to Any Hospital	(5) Crisis Center	(6) Psych Hospital	(7) Emergency Dept with No Psych Unit	(8) Emergency Dept with Psych Unit
Panel A: IV estimates								
MCR response	0.6306	1.4012***	-0.2300	-0.0221	0.0263	0.0420**	-0.0657	-0.0248
	(0.3969)	(0.5079)	(0.1703)	(0.1506)	(0.0510)	(0.0183)	(0.0925)	(0.1173)
Obs	147,614	147,674	147,678	147,678	147,678	147,678	147,678	147,678
Weak id KP Wald F	88.3931	88.3190	88.3140	88.3140	88.3140	88.3140	88.3140	88.3140
Patrol officer mean	7.7166	31.8967	0.7788	0.1574	0.0124	0.0021	0.0575	0.0854
Complier mean under patrol	5.1015	17.6025	0.6759	0.3537	0.0524	-0.0362	0.1615	0.1760
Panel B: OLS Estimates								
MCR response	0.2255***	0.8882***	-0.2462***	0.1205***	0.0554***	0.0009	0.0041	0.1809***
	(0.0102)	(0.0115)	(0.0049)	(0.0044)	(0.0024)	(0.0006)	(0.0026)	(0.0045)
Obs	147,614	147,674	147,678	147,678	147,678	147,678	147,678	147,678
Patrol officer mean	7.7166	31.8967	0.7788	0.1574	0.0134	0.0023	0.0740	0.1189

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation (Panel A) or ordinary least squares estimation (Panel B). Patrol officer mean is the mean outcome for calls handled by regular patrol officers. Complier mean under patrol is the estimated mean outcome for compliers under patrol response (Frandsen *et al.*, 2023), and can fall outside the outcome's range due to sampling noise. In columns 1–2, both are geometric means in minutes. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 5: Downside Risks of Harm (OLS and IV)

	(1) Involuntary Transport	(2) Arrested	(3) Jailed	(4) Force Used
Panel A: IV estimates				
MCR response	0.2142**	0.0760	0.0361	0.0212
	(0.0987)	(0.0503)	(0.0448)	(0.0162)
Obs	147,678	147,678	147,678	147,678
Weak id KP Wald F	88.3140	88.3140	88.3140	88.3140
Patrol officer mean	0.0540	0.0157	0.0125	0.0018
Complier mean under patrol	-0.0247	-0.0356	0.0053	-0.0215
Panel B: OLS Estimates				
MCR response	0.1258***	0.0060***	0.0049***	0.0027***
	(0.0037)	(0.0014)	(0.0013)	(0.0006)
Obs	147,678	147,678	147,678	147,678
Patrol officer mean	0.0540	0.0157	0.0125	0.0018

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation (Panel A) or ordinary least squares estimation (Panel B). Patrol officer mean is the mean outcome for calls handled by regular patrol officers. Complier mean under patrol is the estimated mean outcome for compliers under patrol response (Frandsen *et al.*, 2023), and can fall outside the outcome's range due to sampling noise. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 6: Digitized Use of Force Data

	Digitized 2023 data			Digitized 2024 data		
	(1) Force Used Any	(2) Force Used Low-level (0-1)	(3) Force Used High-level (2-8)	(4) Force Used Any	(5) Force Used Low-level (0-1)	(6) Force Used High-level (2-8)
MCR response	0.1268	0.0888	0.0380	-0.1202*	-0.1119*	-0.0084
	(0.0995)	(0.0847)	(0.0423)	(0.0671)	(0.0604)	(0.0278)
Obs	22,421	22,421	22,421	21,463	21,463	21,463
Weak id KP Wald F	6.3899	6.3899	6.3899	24.4052	24.4052	24.4052
Patrol officer mean	0.0050	0.0038	0.0012	0.0045	0.0032	0.0013
Complier mean under patrol	-0.0716	-0.0495	-0.0221	0.0900	0.0723	0.0177

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Patrol officer mean is the mean outcome for calls handled by regular patrol officers. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 7: Time Spent, Healthcare Access, and Downside Risk of Harm by Call Type

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: Suicide related calls							
MCR response	0.7963	1.2973**	-0.2225	0.2090	0.0875	-0.0266	-0.0124
	(0.5115)	(0.5718)	(0.2656)	(0.2444)	(0.1626)	(0.0579)	(0.0226)
Obs	47698	47710	47711	47711	47711	47711	47711
Weak id KP Wald F	45.4197	45.3795	45.3779	45.3779	45.3779	45.3779	45.3779
Patrol officer mean	11.4184	61.1006	0.6522	0.2547	0.0866	0.0109	0.0018
Panel B: Mental health calls							
MCR response	0.5436	1.4282*	-0.2068	-0.2000	0.2903**	0.0800	0.0456*
	(0.5749)	(0.7597)	(0.2148)	(0.1899)	(0.1238)	(0.0656)	(0.0233)
Obs	99916	99964	99967	99967	99967	99967	99967
Weak id KP Wald F	45.2754	45.2379	45.2373	45.2373	45.2373	45.2373	45.2373
Patrol officer mean	11.2514	50.0122	0.8391	0.1111	0.0384	0.0133	0.0017

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 8: Time Spent, Healthcare Access, and Downside Risk of Harm by Call Severity

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
			Panel A: Low Severity				
MCR response	1.0387	1.7415	-0.2754	0.1466	0.1590	-0.0441	-0.0006
	(0.9345)	(1.2272)	(0.3302)	(0.2969)	(0.1610)	(0.0789)	(0.0156)
Obs	43121	43138	43138	43138	43138	43138	43138
Weak id KP Wald F	18.5590	18.4301	18.4301	18.4301	18.4301	18.4301	18.4301
Dep Mean	11.2772	43.4667	0.8435	0.1231	0.0280	0.0070	0.0004
			Panel B: Medium Severity				
MCR response	0.5600	0.8787	0.1091	-0.3412	0.2374*	0.0462	0.0385**
	(0.5959)	(0.7551)	(0.2606)	(0.2368)	(0.1436)	(0.0604)	(0.0181)
Obs	56605	56618	56620	56620	56620	56620	56620
Weak id KP Wald F	41.8177	42.2197	42.2161	42.2161	42.2161	42.2161	42.2161
Dep Mean	10.8372	50.4332	0.7923	0.1514	0.0486	0.0101	0.0011
			Panel C: High Severity				
MCR response	0.3085	1.0211	-0.3686	0.1182	0.1282	0.0605	0.0078
	(0.6341)	(0.7432)	(0.3111)	(0.2718)	(0.1999)	(0.0949)	(0.0409)
Obs	47642	47672	47674	47674	47674	47674	47674
Weak id KP Wald F	28.5780	28.3266	28.3251	28.3251	28.3251	28.3251	28.3251
Dep Mean	11.8908	66.7679	0.7027	0.1964	0.0844	0.0206	0.0038

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 9: Time Spent, Healthcare Access, and Downside Risk of Harm Impact by Programmatic Era

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: 2018-2020							
MCR response	0.6746	2.0505***	-0.0550	-0.2098	0.2473*	0.0339	0.0255
	(0.5258)	(0.6941)	(0.2401)	(0.2175)	(0.1341)	(0.0609)	(0.0243)
Obs	57707	57709	57712	57712	57712	57712	57712
Weak id KP Wald F	47.3032	47.2721	47.2609	47.2609	47.2609	47.2609	47.2609
Dep Mean	9.7653	52.3953	0.7744	0.1658	0.0497	0.0122	0.0018
Panel B: 2021 only							
MCR response	0.3740	-0.0186	-0.1413	0.4465	-0.2982	-0.0302	0.0014
	(0.7134)	(0.9574)	(0.3028)	(0.2733)	(0.1925)	(0.0723)	(0.0278)
Obs	23151	23160	23160	23160	23160	23160	23160
Weak id KP Wald F	29.2237	29.0908	29.0908	29.0908	29.0908	29.0908	29.0908
Dep Mean	12.3594	53.3103	0.7827	0.1607	0.0490	0.0105	0.0014
Panel C: 2022-2024							
MCR response	1.1364*	1.6493**	-0.4412*	-0.0282	0.3873**	0.0692	0.0211
	(0.6388)	(0.7932)	(0.2614)	(0.2253)	(0.1600)	(0.0727)	(0.0231)
Obs	66756	66805	66806	66806	66806	66806	66806
Weak id KP Wald F	36.5307	36.5409	36.5404	36.5404	36.5404	36.5404	36.5404
Dep Mean	12.2834	54.7272	0.7812	0.1490	0.0594	0.0135	0.0018
Panel D: 2022-2024, reweighted							
MCR response	1.4115*	1.9825**	-0.2712	-0.2151	0.4229**	0.0684	0.0335
	(0.7979)	(1.0083)	(0.2855)	(0.2549)	(0.1693)	(0.0781)	(0.0214)
Obs	66603	66652	66653	66653	66653	66653	66653
Weak id KP Wald F	27.9960	28.0803	28.0800	28.0800	28.0800	28.0800	28.0800
Dep Mean	12.1171	51.0852	0.8092	0.1327	0.0488	0.0118	0.0014

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. Panel D reweights 2022–2024 observations to match the 2018–2020 complier composition across five call type and severity cells (those in Table 10, except pooling low-severity across call types to reduce uncertainty around weights). The dependent mean is also reweighted, and standard errors treat the weights as fixed.

Table 10: Call Characteristics of Compliers by Programmatic Era

		2018–2020		2022–2024	
	All years	Mean	Complier-weighted	Mean	Complier-weighted
	(1)	(2)	(3)	(4)	(5)
Mental health calls					
Low severity	0.221	0.233	0.245	0.209	0.166
	(0.001)	(0.002)	(0.074)	(0.002)	(0.068)
Medium severity	0.254	0.259	0.249	0.246	0.179
	(0.001)	(0.002)	(0.070)	(0.002)	(0.076)
High severity	0.201	0.167	0.164	0.238	0.195
	(0.001)	(0.002)	(0.058)	(0.002)	(0.089)
Suicide-related calls					
Low severity	0.071	0.084	0.054	0.057	0.014
	(0.001)	(0.001)	(0.043)	(0.001)	(0.044)
Medium severity	0.130	0.148	0.155	0.111	0.132
	(0.001)	(0.001)	(0.054)	(0.001)	(0.062)
High severity	0.122	0.109	0.133	0.139	0.314
	(0.001)	(0.001)	(0.051)	(0.001)	(0.078)
Obs	147,432	57,661	57,661	66,653	66,653

Note: Columns (1), (2), and (4) report the share of calls in each call type and severity category in the full estimation sample (2018-2024), in 2018-2020, and in 2022-2024, respectively. Columns (3) and (5) report complier-weighted shares, i.e., shares weighted by the implicit weights of the 2SLS estimates, following Frandsen *et al.* (2023). Each is the coefficient on MCR response from a 2SLS regression of the category indicator multiplied by MCR response on MCR response, instrumented with recent citywide mental illness dispatches and including the time-varying covariates in Equation 1, estimated separately within each era. Bootstrapped standard errors (500 replications) in parentheses.

Table 11: Time Spent, Healthcare Access, and Downside Risk of Harm Impact by Neighborhood Demographics

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: Low Income, Non-White							
MCR response	1.1678	0.8443	0.1833	-0.5317	0.3014	0.0158	0.0075
	(0.9381)	(1.1532)	(0.4119)	(0.3888)	(0.2249)	(0.1122)	(0.0348)
Obs	44444	44464	44466	44466	44466	44466	44466
Weak id KP Wald F	17.5573	17.3785	17.3733	17.3733	17.3733	17.3733	17.3733
Patrol Officer Mean	10.8519	52.0155	0.7718	0.1676	0.0495	0.0134	0.0018
Panel B: Low Income, White							
MCR response	0.2870	0.4452	0.0279	-0.2322	0.1906	0.0083	0.0226
	(0.7603)	(0.9926)	(0.3176)	(0.2845)	(0.1781)	(0.0856)	(0.0307)
Obs	38690	38704	38705	38705	38705	38705	38705
Weak id KP Wald F	27.2119	27.1433	27.1439	27.1439	27.1439	27.1439	27.1439
Patrol Officer Mean	10.3106	50.5443	0.7865	0.1503	0.0519	0.0142	0.0015
Panel C: Higher Income							
MCR response	0.5878	2.2554***	-0.5712**	0.3438*	0.1823	0.0589	0.0269
	(0.5147)	(0.6898)	(0.2331)	(0.2017)	(0.1368)	(0.0558)	(0.0225)
Obs	64322	64348	64349	64349	64349	64349	64349
Weak id KP Wald F	45.6408	45.8389	45.8376	45.8376	45.8376	45.8376	45.8376
Patrol Officer Mean	12.2301	56.5769	0.7790	0.1546	0.0583	0.0110	0.0019

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 12: Transport Location Impact by Neighborhood Demographics

	Voluntary and Involuntary Transports to:			
	(1) Crisis Center	(2) Psych Hospital	(3) Emergency Dept with No Psych Unit	(4) Emergency Dept with Psych Unit
Panel A: Low Income, Non-White				
MCR response	-0.2461	0.0568	0.1263	-0.1312
	(0.1538)	(0.0434)	(0.2542)	(0.3211)
Obs	44466	44466	44466	44466
Weak id KP Wald F	17.3733	17.3733	17.3733	17.3733
Patrol Mean	0.0159	0.0023	0.0834	0.1132
Mean Distance	6.5516	7.1091	4.7642	8.1298
Panel B: Low Income, White				
MCR response	0.0921	0.0639*	-0.2789	0.0892
	(0.1128)	(0.0357)	(0.1988)	(0.2454)
Obs	38705	38705	38705	38705
Weak id KP Wald F	27.1439	27.1439	27.1439	27.1439
Patrol Mean	0.0161	0.0015	0.0677	0.1140
Mean Distance	4.8938	11.4119	4.3131	6.4666
Panel C: Higher Income				
MCR response	0.1133*	0.0351	-0.0199	0.3867**
	(0.0631)	(0.0281)	(0.1365)	(0.1900)
Obs	64349	64349	64349	64349
Weak id KP Wald F	45.8376	45.8376	45.8376	45.8376
Patrol Mean	0.0101	0.0027	0.0713	0.1259
Mean Distance	9.2805	9.1463	5.2686	8.9491

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Mean distance indicates distance from dispatch location to the free crisis center (Column 1), to the nearest psychiatric hospital (Column 2), to the nearest emergency department without a psychiatric unit (Column 3) or to the nearest emergency department with a psychiatric unit (Column 4). Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 13: Time Spent, Healthcare Access, and Potential Risk of Harm Impact by Length of Participation in MCR Program

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: Time spent in MCR program ($\leq$ 2 year)							
MCR response	1.0293*	1.7230**	-0.1466	-0.1564	0.2747*	0.0211	0.0237
	(0.6256)	(0.7965)	(0.2641)	(0.2346)	(0.1508)	(0.0694)	(0.0246)
Obs	142739	142793	142797	142797	142797	142797	142797
Weak id KP Wald F	58.6522	58.6315	58.6611	58.6611	58.6611	58.6611	58.6611
Patrol officer mean	11.3053	53.5887	0.7788	0.1574	0.0540	0.0125	0.0018
MCR officer tenure in CPD	16.5386	16.5384	16.5384	16.5384	16.5384	16.5384	16.5384
Panel B: Time spent in MCR program ($>$ 2 years)							
MCR response	1.1420	2.4199*	-0.0766	-0.3168	0.3268	0.0701	0.0753*
	(0.9922)	(1.2947)	(0.4217)	(0.3780)	(0.2385)	(0.1120)	(0.0412)
Obs	141263	141320	141324	141324	141324	141324	141324
Weak id KP Wald F	31.5607	31.5194	31.4876	31.4876	31.4876	31.4876	31.4876
Patrol officer mean	11.3053	53.5887	0.7788	0.1574	0.0540	0.0125	0.0018
MCR officer tenure in CPD	22.7704	22.7714	22.7714	22.7714	22.7714	22.7714	22.7714

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Time spent in MCR program is calculated as the difference between the date when the senior-most MCR officer was hired into the MCR program and the date of the incident. In 87 percent of MCR responses, only 1 MCR team responds, making the "senior-most MCR officer" responding the only MCR officer responding. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 14: Effect of MCR Response on Resource Utilization

	(1) # Responder Units	(2) # Responders	(3) Responder Hours (log hours)	(4) Patrol Officer Hours (log hours)	(5) Responder Cost (log USD)
MCR response	1.4951**	1.8965**	1.5447***	0.6201	1.5829***
	(0.6604)	(0.7502)	(0.5286)	(0.6335)	(0.5281)
Obs	144056	144056	144056	141900	144056
Weak id KP Wald F	87.4676	87.4676	87.4676	73.4918	87.4676
Patrol officer mean	2.1878	2.6334	1.0717	1.0714	48.6063
Complier mean under patrol	1.5742	2.6286	0.7601	0.7056	34.9433

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Patrol officer mean is the mean outcome for calls handled by regular patrol officers. Complier mean under patrol is the estimated mean outcome for compliers under patrol response (Frandsen *et al.*, 2023), and can fall outside the outcome's range due to sampling noise. In columns 3–5, both are geometric means in hours or USD. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the **?** critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table 15: Repeat Calls Regarding Behavioral Health

	Any repeat BH call within:			
	(1)	(2)	(3)	(4)
	30 days	180 days	# of repeat BH calls	Any repeat with hospitalization
MCR response	-0.0548	0.2815	-0.5592	-0.0011
	(0.2082)	(0.2602)	(2.4784)	(0.1690)
Obs	107,670	100,777	100,777	100,777
Weak id KP Wald F	50.6357	46.4932	46.4932	46.4932
Patrol officer mean	0.1791	0.2904	1.0871	0.0959
Complier mean under patrol	0.2414	0.0433	2.8978	0.1538

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

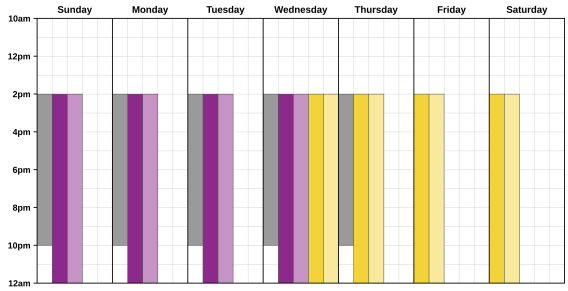
Table 16: Repeat Calls Not Regarding Behavioral Health

	Any repeat non-BH call within:		
	(1)	(2)	(3)
	30 days	180 days	# of repeat non-BH calls
MCR response	0.3227	0.2631	0.8868
	(0.2127)	(0.2751)	(1.5790)
Obs	107,670	100,777	100,777
Weak id KP Wald F	50.6357	46.4932	46.4932
Patrol officer mean	0.1806	0.3611	1.0247
Complier mean under patrol	-0.2389	0.0289	-0.0256

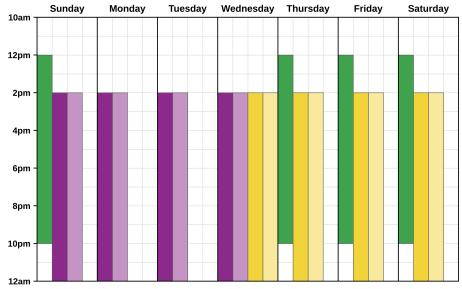
Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Appendix A Supplementary Tables and Figures

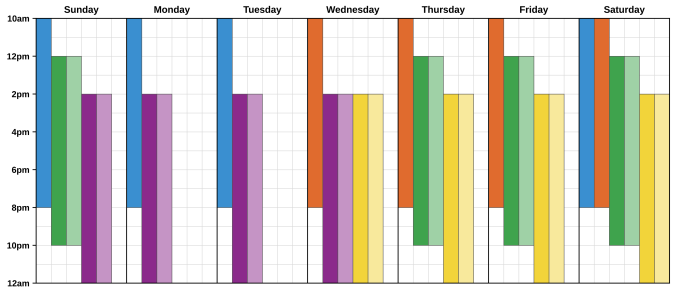
Figure A1: MCR Shifts Over the Sample Period.



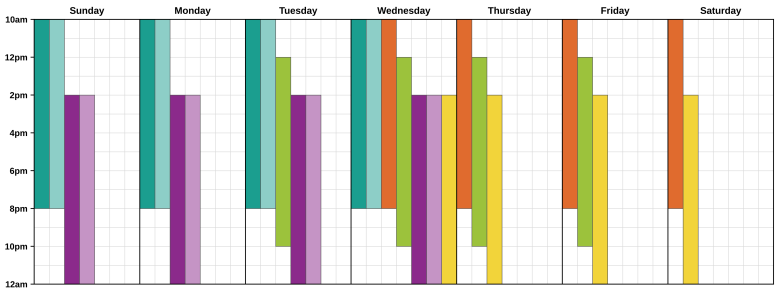
(a) June 3, 2018 – December 15, 2018



(b) December 16, 2018 – August 10, 2019

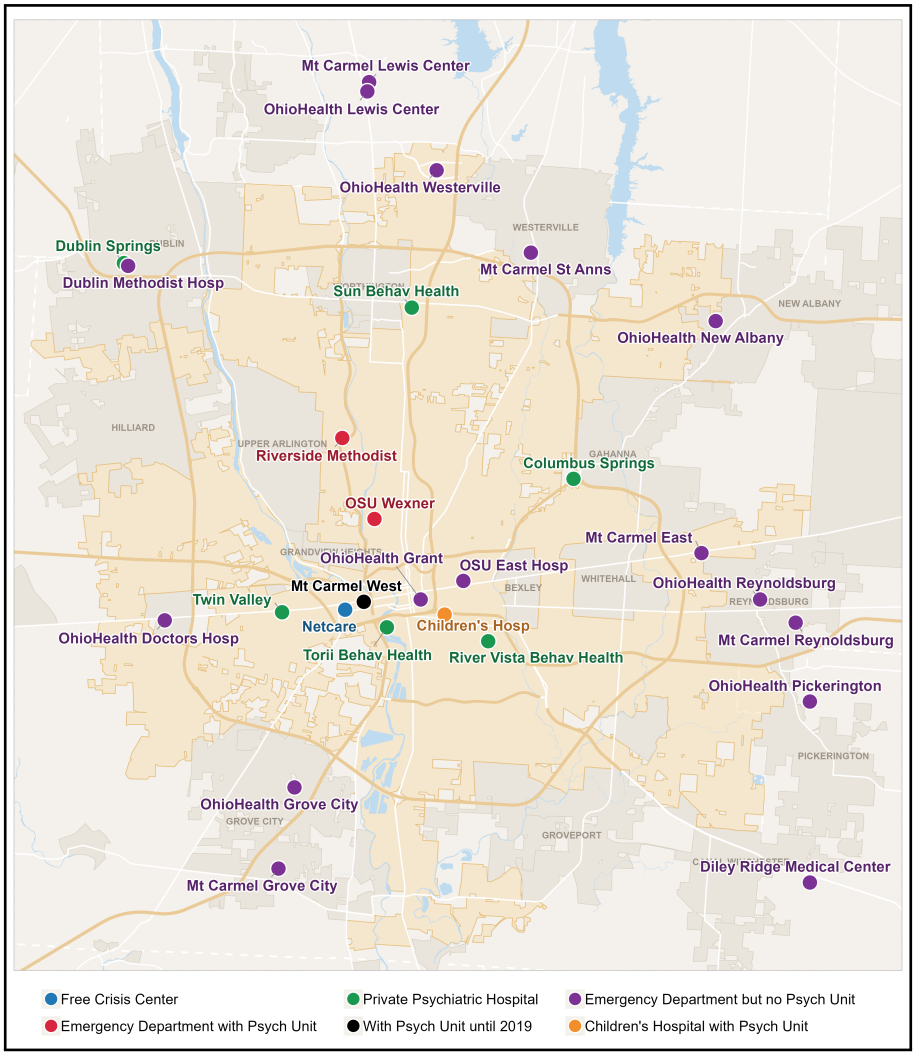


(c) August 11, 2019 – August 20, 2022



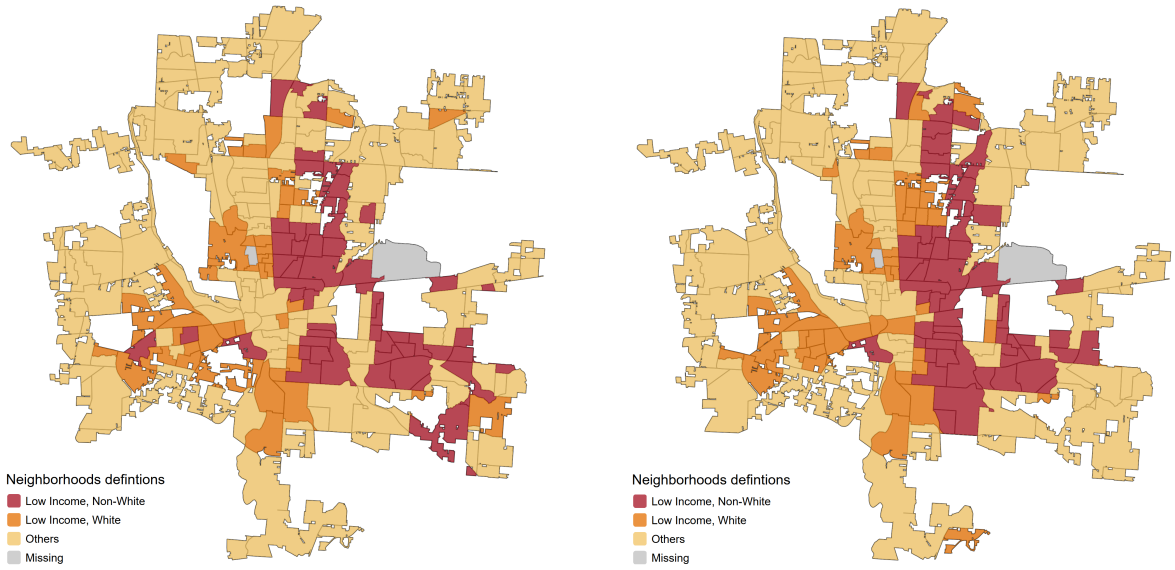
(d) August 21, 2022 – December 31, 2024

Figure A2: Healthcare Facilities in Columbus



This map includes the primary healthcare facilities to which police transports patients. Others are used, but rarely.

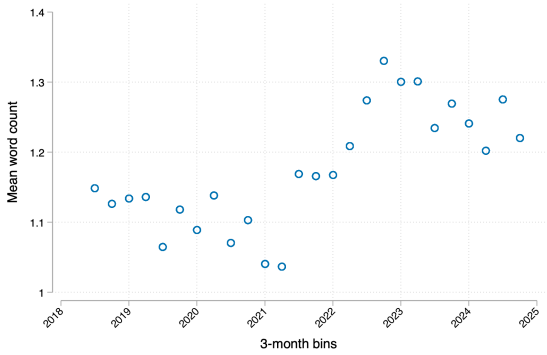
Figure A3: Columbus Neighborhood Categories



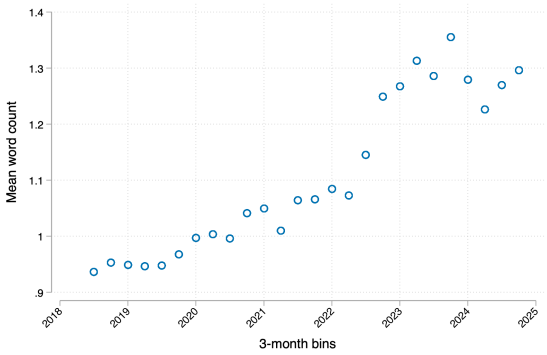
(a) 2020-2024 ACS 5-year estimates **(b)** 2015-2019 ACS 5-year estimates

Note: U.S. Census tracts are grouped into three neighborhood categories based on socioeconomic characteristics drawn from the ACS five-year estimates. Panel (a) shows neighborhood categories based on the 2020-2024 ACS five-year estimates, which were used for dispatch data from 2020-2024, whereas panel (b) shows neighborhood categories based on the 2015-2019 ACS five-year estimates, which were used for dispatch data from 2018-2019. Using the 2020-2024 ACS five-year estimates, census tracts with a non-white population of at least 50% and a median household income of $60,000 or less were classified as lower-income minority; tracts with a non-white population of less than 50% and a median household income of $60,000 or less were classified as lower-income white; all other tracts were classified as higher-income. Using the 2015-2019 ACS five-year estimates, the same classification was applied with a median household income cutoff of $50,000. Income cutoffs were based on 200% of the federal poverty line for a family of four as of 2024 and 2019, respectively.

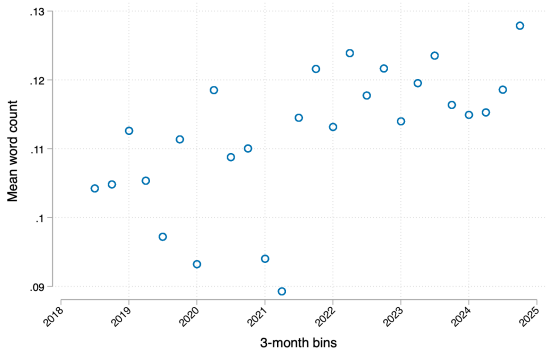
Figure A4: Word Count Trends for Sample Dispatches



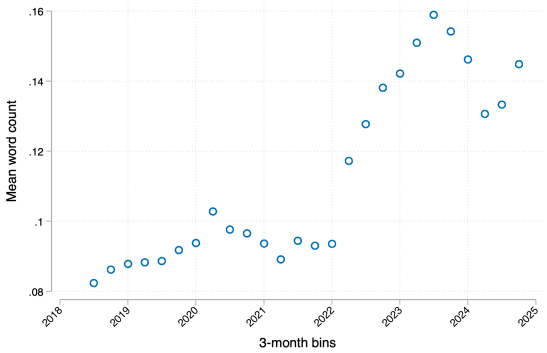
(a) Risk of Harm Words



(b) Psychiatric Acuity Words



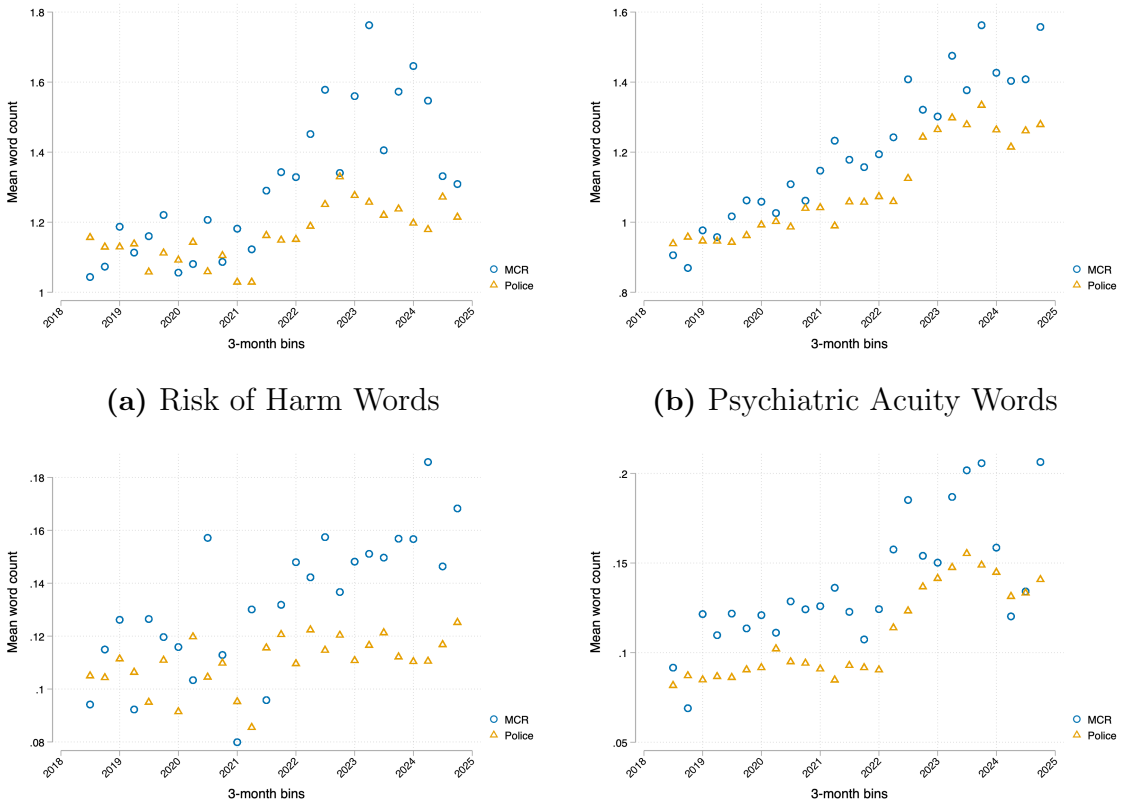
(c) Mentions of Weapons



(d) Mentions of Schizophrenia

Note: Graphics demonstrate the average number of words in the dispatch notes field for calls in the binned period related to: risk of harm (or violence), psychiatric acuity, weapons in particular, or schizophrenia in particular.

Figure A5: Word Count Trends for MCR and Police Dispatches Separately



(a) Risk of Harm Words

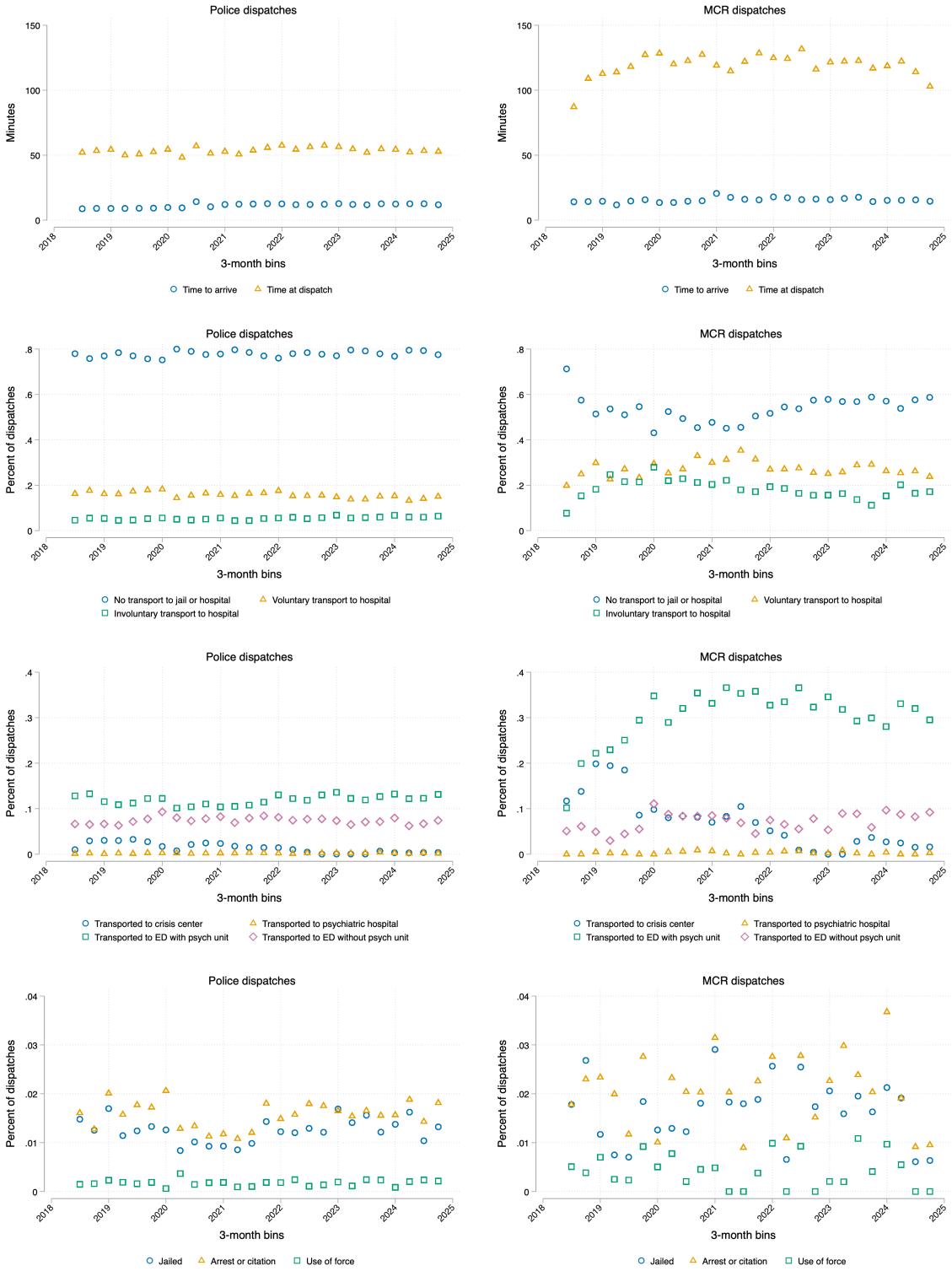
(b) Psychiatric Acuity Words

(c) Mentions of Weapons

(d) Mentions of Schizophrenia

Note: Graphics demonstrate the average number of words in the dispatch notes field for calls in the binned period, and for the responder category indicated, related to: risk of harm (or violence), psychiatric acuity, weapons in particular, or schizophrenia in particular.

Figure A6: Time Trends in Primary Outcomes



Note: On the left are time trends for police dispatches; on the right, time trends for MCR dispatches. The top row shows binned averages for minutes between call and arrival ("time to arrival") and minutes between call and dispatch closing ("time at dispatch"). The second row shows the percent of dispatches, in each bin, that are transported neither to a jail nor to a hospital; who are voluntarily transported to a hospital; or who are involuntarily transported to a hospital. The third row shows the percent of dispatches in each bin that are voluntarily transported to the free crisis center; to a private psychiatric hospital; to an emergency department with a psychiatric unit; or to an emergency department without a psychiatric unit. The bottom row shows the percent of dispatches in each bin that lead to jailing; to an arrest; or to use of force (as per the police Internal Affairs database). The entire sample period is shown, except that July 2018 is dropped for visual ease, being outside a 3-month bin.

Table A2: Categorical Designation of Calls by Dispatcher-Assigned Police Code

Remarks based	Tencode based				
	Intoxicated Person	Disturb Mental	Overdose Accidental	Suicide Attempt	Other Tencodes
Substance Use	3,534	1,598	1,054	1,202	101,688
Mental Illness	66	30,828	123	2,624	46,685
Overdose	8	72	14,374	212	3,628
Suicide	11	7,803	207	24,613	9,346
All mental health calls	3,619	40,301	15,758	28,651	161,347
N	4,440	58,639	20,055	34,505	2,127,117

NOTE: This table shows behavioral health call classifications based on both the assigned police code and call remarks, for nearly 2.8 million calls successfully dispatched within Columbus police precincts between June 2018 and December 2024. Column 2 reports classifications based on call remarks for calls originally coded as either "disturbance - mental" (10-16B) or "indecent exposure" (10-44B) by dispatch employees.

Figure A7: Daily Trends in Instrument Value

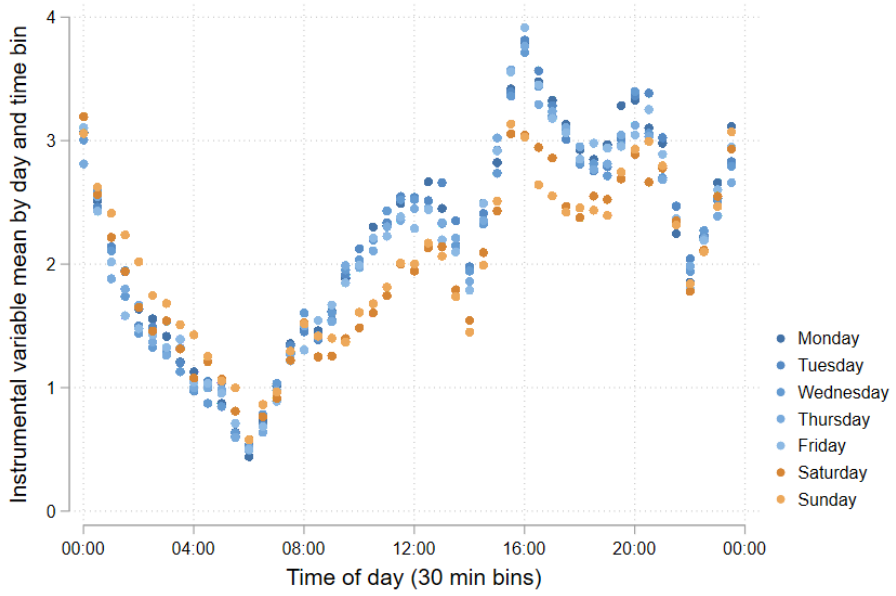


Table A1: Police Dispatch Record - Mock data

Event No.	Event Dispatched	Event Priority	Police Code	Description	Event Remarks	Units on Event	Disposition Codes	Transport
P180004522	20181101001834	2	47A	Suicide Attempt	15yo caller feeling suicidal and depressed..is bipolar and off meds	MCR4	4,2	MCR4 – Changed location to @drs west hospital
P190033324	20190212093513	3	16B	Disturbance/ Mental	male client banging head on the wall, screaming, out of control	MCR2 MCR5	4, 2	MCR5 – Changed location to @netc ∗||∗ MCR5 – Changed location to @riverside hosp
P230577650	20230504172301	3	19	Intoxicated Person	extremely intoxicated male is staggering, moving in and out of traffic...	34B MCR3	2	–
P230976388	20231018142736	2	34O	Overdose - Accidental	female possibly overdosing turning blue breathing slightly...	9190E 53B MCR1	4,2,1	53B – Changed location to osu main
P240011874	20241223222048	2	34C	Check on well being	Callers friend is in distress, has attempted suicide in the past by overdosing	MCR1 MCR2	2, 4	MCR2 – Changed location to @netcare

Note: This is a mock dataset providing a sample view of the data, including just a subset of data fields from the dispatch records. In the "Disposition Codes" field, code 1 indicates "Report Completed and/or Citation Issued", code 2 indicates "Contact made / Party advised or no report needed", code 3 indicates "Arrest Made", and disposition code 4 stands for "Non-Arrest Situation or Errand Complete".

Table A3: First Stage Regression across instruments with varying time window

	(1) MCR responded	(2) MCR responded	(3) MCR responded	(4) MCR responded	(5) MCR responded
# of MH dispatches in 30 min prior to dispatch i from OTHER precincts	-0.0046***				
	(0.0006)				
# of MH dispatches in 40 min prior to dispatch i from OTHER precincts		-0.0042***			
		(0.0005)			
# of MH dispatches in 50 min prior to dispatch i from OTHER precincts			-0.0041***		
			(0.0004)		
# of MH dispatches in 60 min prior to dispatch i from OTHER precincts				-0.0034***	
				(0.0004)	
# of MH dispatches in 70 min prior to dispatch i from OTHER precincts					-0.0032***
					(0.0004)
Obs	147,678	147,678	147,678	147,678	147,678
Weak id KP Wald F	67.2815	74.4054	88.3140	71.2285	73.4664
Underid KP LM stat	67.1408	74.2579	88.1331	71.1358	73.3775
Underid LM stat p-val	0.0000	0.0000	0.0000	0.0000	0.0000
Dep Mean	0.0750	0.0750	0.0750	0.0750	0.0750
Weekend × Adapted Cubic Spline	Y	Y	Y	Y	Y
Annual 1st-order Fourier	Y	Y	Y	Y	Y
Time, Time^2	Y	Y	Y	Y	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates are obtained after controlling for diurnal trends, annual trends, and time trends via our primary specification (Equations 1-3). In addition, Column 2 controls for dispatcher-assigned codes (see Table A4) and Column 3 controls for holidays, moon phases and illumination and weather (see Table A5). We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. We also report the Kleibergen-Paap rk LM heteroskedasticity-robust underidentification test statistic with the corresponding p-value reported below.

Table A4: Instrument Balance on Dispatcher-Assigned Codes

	(1) MCR	(2) IV
DISTURBANCE	-0.0199***	0.0299
	(0.0031)	(0.0219)
DISTURBANCE / MENTAL	0.0404***	0.0157
	(0.0030)	(0.0182)
DOMESTIC DISPUTE	-0.0198***	0.0149
	(0.0039)	(0.0304)
DOMESTIC VIOLENCE	-0.0386***	-0.0226
	(0.0039)	(0.0278)
GUN/KNIFE	0.0153***	-0.0171
	(0.0043)	(0.0256)
INDECENT EXPOSURE	-0.0573***	0.0412
	(0.0029)	(0.0305)
INTOXICATED PERSON	-0.0336***	0.1375
	(0.0078)	(0.1741)
OTHER	-0.0084**	0.0100
	(0.0033)	(0.0215)
OVERDOSE - ACCIDENTAL	-0.0461***	0.0387
	(0.0081)	(0.0875)
SUICIDE ATTEMPT	0.0099***	0.0208
	(0.0035)	(0.0215)
Call coded as urgent (priority 1,2)	0.0142***	-0.0144
	(0.0020)	(0.0116)
Obs	147678	147678
Dep Mean	0.0750	2.4010
Rsq	0.0620	0.1731
Joint test F-stat	358.2130	1.5610
Joint test p-val	0.0000	0.1030
Weekend × Adapted Cubic Spline	Y	Y
Annual 1st-order Fourier	Y	Y
Time, Time^2	Y	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates are obtained after controlling for diurnal trends, annual trends, and time trends – via our primary specification in Columns 1 and 2, and via weekend-hour, month, and year fixed effects in Column 3. Controls are identical in Columns 1 and 4 except that we use a 6th-order rather than 2nd-order truncated Fourier series for capturing diurnal trends in Column 4.

Table A5: Instrument Balance on Holidays, Ohio State Game Days, Moon Phases and Illumination, and Weather

	(1) MCR	(2) IV
Holiday (0/1)	0.0061	-0.2279***
	(0.0042)	(0.0240)
Ohio State Game Days (0/1)	-0.0065*	-0.0826***
	(0.0037)	(0.0227)
Moon Illumination	0.0011	-0.0690***
	(0.0044)	(0.0264)
Phase 1: New Moon (0/1)	-0.0025	-0.0542***
	(0.0027)	(0.0166)
Phase 2: First Quarter (0/1)	0.0035*	-0.0254**
	(0.0019)	(0.0116)
Phase 3: Full Moon (0/1)	0.0003	-0.0115
	(0.0028)	(0.0165)
Avg Daily Wind Speed (mph)	-0.0002	-0.0096***
	(0.0002)	(0.0013)
Daily Precipitation (inches)	-0.0081***	-0.0485***
	(0.0022)	(0.0139)
Daily Max Temperature (F)	-0.0001	0.0052***
	(0.0001)	(0.0006)
Daily Min Temperature (F)	0.0001	0.0036***
	(0.0001)	(0.0007)
Obs	147568	147568
Dep Mean	0.0750	2.4010
Rsq	0.0520	0.1757
Joint test F-stat	1.8750	57.8370
Joint test p-val	0.0590	0.0000
Weekend × Adapted Cubic Spline	Y	Y
Annual 1st-order Fourier	Y	Y
Time, Time^2	Y	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates are obtained after controlling for diurnal trends, annual trends, and time trends – via our primary specification in Columns 1 and 2, and via weekend-hour, month, and year fixed effects in Column 3. Controls are identical in Columns 1 and 4 except that we use a 6th-order rather than 2nd-order truncated Fourier series for capturing diurnal trends in Column 4.

Table A6: First Stage Regression with Unbalanced Time-Varying Events

	(1) MCR responded	(2) MCR responded
# MH dispatches in 50 min prior to dispatch i from OTHER precincts	-0.0041***	-0.0041***
	(0.0004)	(0.0004)
Obs	147,678	147,568
Weak id KP Wald F	88.3140	88.6534
Underid KP LM stat	88.1331	88.4741
Underid LM stat p-val	0.0000	0.0000
Dep Mean	0.0750	0.0750
Weekend × Weekend × Adapted Cubic Spline	Y	Y
Annual 1st-order Fourier	Y	Y
Time, Time^2	Y	Y
Holidays, game, moon, and weather controls	N	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 2 controls for holidays, Ohio State game days, moon phases and illumination and weather (see Table A5). We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. We also report the Kleibergen-Paap rk LM heteroskedasticity-robust underidentification test statistic with the corresponding p-value reported below.

Table A7: Right Response Unit First Stage Regression

	(1) Right Response All	(2) Right Response Low-Medium Severity	(3) Right Response High Severity	(4) Right Response Critical Severity
# MH dispatches in 50 min prior to dispatch i from OTHER precincts	-0.0002	0.0009	-0.0005	-0.0005
	(0.0004)	(0.0007)	(0.0006)	(0.0008)
Obs	147,678	37,264	67,542	42,624
Weak id KP Wald F	0.2049	1.3984	0.7564	0.4720
Underid KP LM stat	0.2050	1.3993	0.7567	0.4723
Underid LM stat p-val	0.6507	0.2368	0.3843	0.4919
Dep Mean	0.0660	0.0590	0.0630	0.0750
Weekend × Adapted Cubic Spline	Y	Y	Y	Y
Annual 1st-order Fourier	Y	Y	Y	Y
Time, Time^2	Y	Y	Y	Y

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Column 2 controls for holidays, moon phases and illumination and weather (see Table A5). We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. We also report the Kleibergen-Paap rk LM heteroskedasticity-robust underidentification test statistic with the corresponding p-value reported below.

Table A8: Monotonicity: First Stage by Call and Dispatch Characteristics

	(1) 911 Call	(2) From Netcare	(3) Passerby	(4) Anonymous	(5) Black	(6) White
MCR Instrument	-0.0033***	-0.0076***	-0.0082*	-0.0042***	-0.0030***	-0.0037***
	(0.0006)	(0.0023)	(0.0045)	(0.0014)	(0.0009)	(0.0009)
Obs	70,184	7,677	1,694	10,723	36,505	39,871
KP Wald F	26.7728	11.4592	3.3657	9.0263	12.2742	19.0025
	(7) Hispanic	(8) Female	(9) Male	(10) Both Genders	(11) Kid	(12) Family
MCR Instrument	0.0026	-0.0040***	-0.0037***	-0.0043***	-0.0034***	-0.0042***
	(0.0036)	(0.0009)	(0.0007)	(0.0007)	(0.0012)	(0.0007)
Obs	2,939	40,958	49,183	51,813	19,452	56,310
KP Wald F	0.5464	21.8291	26.1561	34.3040	7.8799	32.9826
	(13) Partner	(14) Houseless	(15) Weapon	(16) Violent	(17) Offmeds	(18) Crazy
MCR Instrument	-0.0033***	-0.0067**	-0.0013	-0.0027	-0.0046*	-0.0028
	(0.0011)	(0.0027)	(0.0015)	(0.0018)	(0.0024)	(0.0021)
Obs	20,220	3,346	16,514	9,967	6,211	4,107
KP Wald F	8.4619	6.1502	0.7343	2.4218	3.4766	1.8447
	(19) Schizophrenic	(20) Depressed	(20) Well Being Check	(20) Disturbance	(20) Disturbance Mental	(20) Domestic Dispute
MCR Instrument	-0.0036**	-0.0072***	-0.0026	-0.0021*	-0.0054***	0.0009
	(0.0015)	(0.0018)	(0.0017)	(0.0011)	(0.0008)	(0.0018)
Obs	16,558	8,599	9,083	12,729	52,793	3,719
KP Wald F	6.0186	15.4995	2.2935	3.8362	41.5742	0.2748
	(20) Domestic Violence	(20) Person with Gun/Knife	(20) Indecent Exposure	(20) Suicide	(20) Other Police Codes	(20) Urgent Calls
MCR Instrument	-0.0018	-0.0040**	-0.0004	-0.0065***	-0.0007	-0.0047***
	(0.0015)	(0.0019)	(0.0006)	(0.0009)	(0.0011)	(0.0006)
Obs	6,121	8,844	3,878	33,043	17,063	80,894
KP Wald F	1.5186	4.4237	0.4244	47.3273	0.3614	61.2454

Table A9: Monotonicity: First Stage by Call Type, Clinician Era, Call Severity, and Neighborhood Type

	Call Type		Clinician Era		Call Severity				Neighborhood Type	
	(1) Mental Illness	(2) Suicide	(3) Netcare	(4) CPH Clinicians	(5) Low Severity	(6) Medium Severity	(7) High Severity	(8) Critical Severity	(9) Low-Income Non-White	(10) Low-Income White
MCR Instrument	-0.0036***	-0.0052***	-0.0047***	-0.0040***	-0.0035***	-0.0043***	-0.0043***	-0.0032***	-0.0044***	-0.0046***
	(0.0005)	(0.0008)	(0.0006)	(0.0007)	(0.0008)	(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0007)
Obs	99,967	47,711	80,929	66,749	43,138	56,620	47,674	44,466	38,705	64,349
KP Wald F	45.2373	45.3779	65.5070	36.3691	18.4301	42.2161	28.3251	17.3733	27.1439	45.8376

NOTE: Census tracts with a non-white population of at least 50% and a median household income of $60,000 or less were classified as lower-income minority. Those with a non-white population of less than 50% and a median household income of $60,000 or less were classified as lower-income white. All other tracts were classified as higher-income neighborhoods. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. We also report the Kleibergen-Paap rk LM heteroskedasticity-robust underidentification test statistic with the corresponding p-value reported below.

Table A10: MCR first stage compliers and non-compliers

	By Dispatch hour		By Dispatcher Assigned Priority	
	(1) MCR responded	(2) MCR responded	(3) MCR responded	(4) MCR responded
# MH dispatches in 50 min prior to dispatch i from OTHER precincts	0.0002	-0.0052***	-0.0011	-0.0042***
	(0.0002)	(0.0005)	(0.0023)	(0.0004)
Obs	40,641	107,037	1,933	145,745
Weak id KP Wald F	0.5943	91.4041	0.2287	88.5239
Underid KP LM stat	0.5946	91.1757	0.2322	88.3390
Underid LM stat p-val	0.4406	0.0000	0.6299	0.0000
Dep Mean	0.0040	0.1030	0.0200	0.0760
Weekend × Adapted Cubic Spline	Y	Y	Y	Y
Annual 1st-order Fourier	Y	Y	Y	Y
Time, Time^2	Y	Y	Y	Y
Sample	Dispatch hour (0:00-9:59)	Dispatch hour (10:00-23:59)	Low Priority (4,5)	Medium-High Priority (1,2,3)

NOTE: Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. These estimates are obtained after controlling for first and second order fourier series, dispatcher assigned police code categories, as well as day-of-week, month, and year fixed effects. We classify calls assigned eventpriority of 1 or 2 as high priority, those assigned priority 3 as medium priority, and those assigned priority 4 and 5 as low priority. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively. We also report the Kleibergen-Paap rk LM heteroskedasticity-robust underidentification test statistic with the corresponding p-value reported below.

Table A11: Involuntary transport to different healthcare facilities (IV)

	Involuntary Transport to:			
	(1) Crisis Center	(2) Psych Hospital	(3) Emergency Dept with No Psych Unit	(4) Emergency Dept with Psych Unit
MCR response	-0.0025	0.0062	0.0167	0.1988**
	(0.0191)	(0.0072)	(0.0515)	(0.0810)
Obs	147,678	147,678	147,678	147,678
Weak id KP Wald F	88.3140	88.3140	88.3140	88.3140
Patrol officer mean	0.0010	0.0002	0.0165	0.0335

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table A12: Public-order, Drug-related, and Addiction-supporting Charges

		Arrest Charges:	
	(1) Public Order Related	(2) Drug Related	(3) Addiction Supporting
MCR response	0.0167	-0.0165	-0.0050
	(0.0197)	(0.0114)	(0.0087)
Obs	147,522	147,522	147,522
Weak id KP Wald F	88.0844	88.0844	88.0844
Patrol officer mean	0.0024	0.0007	0.0004

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Serious Charges as Classified by NIBRS Codes

	Arrest Charges:			
	(1) Crime against Person (Serious)	(2) Crime against Property (Serious)	(3) Crime against Society (Serious)	(4) Other Offenses (Low-Level)
MCR response	0.0437	0.0187	-0.0076	-0.0006
	(0.0372)	(0.0167)	(0.0116)	(0.0173)
Obs	147,522	147,522	147,522	147,522
Weak id KP Wald F	88.0844	88.0844	88.0844	88.0844
Patrol officer mean	0.0082	0.0017	0.0007	0.0019

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Time Spent and Healthcare Access with Time Cell Fixed Effects (Blandhol *et al.*, 2026)

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport to Any Hospital	Voluntary Transport to: (5) Crisis Center	(6) Psych Hospital	(7) Emergency Dept with No Psych Unit	(8) Emergency Dept with Psych Unit
Panel A: IV estimates								
MCR response	0.3595	1.4343***	-0.2573	0.0198	0.0324	0.0434**	-0.0551	-0.0009
	(0.3920)	(0.5060)	(0.1696)	(0.1495)	(0.0507)	(0.0181)	(0.0922)	(0.1165)
Obs	147,614	147,674	147,678	147,678	147,678	147,678	147,678	147,678
Weak id KP Wald F	89.5832	89.5352	89.5256	89.5256	89.5256	89.5256	89.5256	89.5256
Officer mean	11.3053	53.5887	0.7788	0.1574	0.0124	0.0021	0.0575	0.0854

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation, with additional fixed effects added to the equation: weekend-hour fixed effects and month and year fixed effects. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table A15: Downside Risks of Harm with Time Cell Fixed Effects (Blandhol *et al.*, 2026)

	(1) Involuntary Transport	(2) Arrested	(3) Jailed	(4) Force Used
Panel A: IV estimates				
MCR response	0.2067**	0.0688	0.0286	0.0207
	(0.0983)	(0.0501)	(0.0446)	(0.0163)
Obs	147,678	147,678	147,678	147,678
Weak id KP Wald F	89.5256	89.5256	89.5256	89.5256
Officer mean	0.0540	0.0157	0.0125	0.0018

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation, with additional fixed effects added to the equation: weekend-hour fixed effects and month and year fixed effects. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table A16: Time Spent and Healthcare Access with Weather, Holiday and Football Controls

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport to Any Hospital	Voluntary Transport to: (5) Crisis Center	(6) Psych Hospital	(7) Emergency Dept with No Psych Unit	(8) Emergency Dept with Psych Unit
Panel A: IV estimates								
MCR response	0.5261	1.2567**	-0.1988	-0.0462	0.0255	0.0411**	-0.0777	-0.0352
	(0.3955)	(0.5060)	(0.1702)	(0.1508)	(0.0510)	(0.0183)	(0.0925)	(0.1174)
Obs	147,504	147,564	147,568	147,568	147,568	147,568	147,568	147,568
Weak id KP Wald F	88.7137	88.6586	88.6534	88.6534	88.6534	88.6534	88.6534	88.6534
Officer mean	11.3056	53.5885	0.7788	0.1575	0.0124	0.0021	0.0575	0.0855

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation, with additional controls added to the equation – holidays, Ohio State game days, moon illumination and moon phases, average daily wind speed, daily precipitation, and daily maximum and minimum temperature. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table A17: Downside Risks of Harm with Weather, Holiday and Football Controls

	(1) Involuntary Transport	(2) Arrested	(3) Jailed	(4) Force Used
Panel A: IV estimates				
MCR response	0.2107**	0.0768	0.0317	0.0207
	(0.0986)	(0.0502)	(0.0447)	(0.0161)
Obs	147,568	147,568	147,568	147,568
Weak id KP Wald F	88.6534	88.6534	88.6534	88.6534
Officer mean	0.0540	0.0156	0.0125	0.0018

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation, with additional controls added to the equation – holidays, Ohio State game days, moon illumination and moon phases, average daily wind speed, daily precipitation, and daily maximum and minimum temperature. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table A18: Zero First Stage Placebo: Time Spent and Healthcare Access (Reduced Form Equation in Overnight Calls)

					Voluntary Transport to:			
	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport to Any Hospital	(5) Crisis Center	(6) Psych Hospital	(7) Emergency Dept with No Psych Unit	(8) Emergency Dept with Psych Unit
Instrument	-0.0031	-0.0065	0.0022	-0.0009	-0.0005	-0.0002	-0.0009	0.0007
	(0.0036)	(0.0049)	(0.0016)	(0.0014)	(0.0004)	(0.0002)	(0.0009)	(0.0010)
Obs	40,627	40,637	40,641	40,641	40,641	40,641	40,641	40,641
Officer mean	10.4880	50.4759	0.7787	0.1600	0.0114	0.0018	0.0685	0.0783

Note: The coefficient listed is from an OLS regression of the outcome on the instrument (the number of mental illness and suicide related dispatches in other precincts in the 50 minutes before dispatch), with the temporal controls in Equation 1. The sample is restricted to calls dispatched between midnight and 10 a.m., when MCR teams are off duty and the instrument does not predict MCR response. Under the exclusion restriction, these coefficients should be zero. Officer mean is the outcome mean among calls handled by regular officers. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A19: Zero First Stage Placebo: Downside Risks of Harm (Reduced Form Equation in Overnight Calls)

	(1) Involuntary Transport	(2) Arrested	(3) Jailed	(4) Force Used
Instrument	-0.0010	-0.0005	-0.0002	-0.0001
	(0.0008)	(0.0005)	(0.0004)	(0.0002)
Obs	40,641	40,641	40,641	40,641
Officer mean	0.0493	0.0176	0.0142	0.0021

Note: The coefficient listed is from an OLS regression of the outcome on the instrument (the number of mental illness and suicide related dispatches in other precincts in the 50 minutes before dispatch), with the temporal controls in Equation 1. The sample is restricted to calls dispatched between midnight and 10 a.m., when MCR teams are off duty and the instrument does not predict MCR response. Under the exclusion restriction, these coefficients should be zero. Officer mean is the outcome mean among calls handled by regular officers. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A20: Zero First Stage Placebo: Time Spent and Healthcare Access (Reduced Form Equation in Low-Priority Calls)

					Voluntary Transport to:			
	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport to Any Hospital	(5) Crisis Center	(6) Psych Hospital	(7) Emergency Dept with No Psych Unit	(8) Emergency Dept with Psych Unit
Instrument	-0.0066	0.0371	-0.0002	-0.0012	-0.0002	0.0000	-0.0016	0.0006
	(0.0164)	(0.0289)	(0.0028)	(0.0024)	(0.0007)	(.)	(0.0013)	(0.0019)
Obs	1,933	1,932	1,933	1,933	1,933	1,933	1,933	1,933
Officer mean	12.3633	47.4535	0.9694	0.0248	0.0016	0.0000	0.0069	0.0164

Note: The coefficient listed is from an OLS regression of the outcome on the instrument (the number of mental illness and suicide related dispatches in other precincts in the 50 minutes before dispatch), with the temporal controls in Equation 1. The sample is restricted to low-priority calls (dispatcher-assigned priority 4–5), for which the instrument does not predict MCR response. Under the exclusion restriction, these coefficients should be zero. Officer mean is the outcome mean among calls handled by regular officers. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A21: Zero First Stage Placebo: Downside Risks of Harm (Reduced Form Equation in Low-Priority Calls)

	(1) Involuntary Transport	(2) Arrested	(3) Jailed	(4) Force Used
Instrument	0.0004	0.0001	0.0009	-0.0002
	(0.0013)	(0.0008)	(0.0007)	(0.0008)
Obs	1,933	1,933	1,933	1,933
Officer mean	0.0032	0.0063	0.0026	0.0021

Note: The coefficient listed is from an OLS regression of the outcome on the instrument (the number of mental illness and suicide related dispatches in other precincts in the 50 minutes before dispatch), with the temporal controls in Equation 1. The sample is restricted to low-priority calls (dispatcher-assigned priority 4–5), for which the instrument does not predict MCR response. Under the exclusion restriction, these coefficients should be zero. Officer mean is the outcome mean among calls handled by regular officers. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A22: Test of Strict Monotonicity and Exclusion (Frandsen *et al.*, 2023)

	Slope p-value	Fit p-value	Joint p-value
Left at Home (No Transport to Jail, Hospital)	0.944	0.055	0.111
Voluntary Transport to Any Hospital	0.963	0.094	0.189
Voluntary Transport to Crisis Center	1.000	0.000	0.000
Voluntary Transport to Psychiatric Hospital	1.000	0.002	0.005
Voluntary Transport to Emergency Dept with No Psych Unit	1.000	0.134	0.268
Voluntary Transport to Emergency Dept with a Psych Unit	0.993	0.002	0.005
Involuntarily transported	0.899	0.179	0.358
Arrested	0.791	0.103	0.206
Jailed	0.779	0.089	0.178
Force used	1.000	0.000	0.000
Obs		147,678	

Note: The table reports p-values from the joint test of strict exclusion and pairwise monotonicity proposed by Frandsen *et al.* (2023), implemented with the semiparametric procedure in their online appendix (Stata package `testjfe`). Following their approach for continuous instruments, each value of the instrument (the leave-out count of mental illness dispatches in other precincts in the prior 50 minutes, top-coded at 7, the 99th percentile value) is treated as a separate "judge." Outcomes and MCR assignment are adjusted for the time-varying covariates in Equation 1 by regressing each on the controls and instrument-value indicators and subtracting the estimated control component. Under both assumptions, average outcomes by instrument value must be a continuous function of the instrument-level MCR propensity, with slope bounded by the outcome's support (± 1). The test fits a flexible spline to instrument-level average outcomes and propensities and, following Chyn *et al.* (2025), checks two conditions. The slope p-value tests whether the slope of the fitted function is impossibly large, i.e., exceeds the range of possible treatment effects. The fit p-value tests whether instrument value has significant explanatory power for the outcome after accounting for each instrument value's predicted point on the fitted function, i.e., whether the distance from the observed points to the fitted function is consistent with sampling variation. The joint p-value combines the two components using a Bonferroni correction. Log time outcomes are excluded because they have no bounded support. Each test uses the corresponding 2SLS estimation sample.

Table A23: Time Spent, Healthcare Access, and Downside Risk of Harm by Call Severity for Mental Health Calls Only

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
			Panel A: Low Severity				
MCR response	1.3102	1.1223	-0.2021	0.0867	0.1419	-0.0101	0.0062
	(1.0234)	(1.3363)	(0.2970)	(0.2638)	(0.1419)	(0.0807)	(0.0158)
Obs	32614	32629	32629	32629	32629	32629	32629
Weak id KP Wald F	16.5640	16.4317	16.4317	16.4317	16.4317	16.4317	16.4317
Dep Mean	11.3654	40.0810	0.8934	0.0825	0.0180	0.0072	0.0005
			Panel B: Medium Severity				
MCR response	-0.2771	1.1052	0.2197	-0.7773**	0.4836**	0.1096	0.0703**
	(0.9285)	(1.2080)	(0.3510)	(0.3511)	(0.2018)	(0.0971)	(0.0322)
Obs	37456	37466	37467	37467	37467	37467	37467
Weak id KP Wald F	19.3875	19.7435	19.7444	19.7444	19.7444	19.7444	19.7444
Dep Mean	10.7300	46.7074	0.8552	0.1018	0.0335	0.0111	0.0011
			Panel C: High Severity				
MCR response	0.3541	1.0074	-0.5010	0.0609	0.0915	0.1277	0.0473
	(1.1450)	(1.3681)	(0.5295)	(0.4452)	(0.3292)	(0.1824)	(0.0751)
Obs	29630	29653	29655	29655	29655	29655	29655
Weak id KP Wald F	9.3027	9.1386	9.1378	9.1378	9.1378	9.1378	9.1378
Dep Mean	11.7896	65.4050	0.7575	0.1552	0.0676	0.0230	0.0038

Note:The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A24: Time Spent, Healthcare Access, and Downside Risk of Harm by Call Severity for Suicide Related Calls Only

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
			Panel A: Low Severity				
MCR response	0.1810	5.5943	-1.0196	0.7901	0.2934	-0.2065	-0.0336
	(2.3032)	(4.0265)	(1.2948)	(1.1762)	(0.6376)	(0.2662)	(0.0534)
Obs	10507	10509	10509	10509	10509	10509	10509
Weak id KP Wald F	2.4569	2.4523	2.4523	2.4523	2.4523	2.4523	2.4523
Dep Mean	11.0068	53.8554	0.6904	0.2478	0.0587	0.0063	0.0003
			Panel B: Medium Severity				
MCR response	1.5507**	0.4410	0.0807	0.0966	-0.0777	-0.0300	-0.0021
	(0.7637)	(0.7867)	(0.3802)	(0.3465)	(0.2233)	(0.0684)	(0.0170)
Obs	19149	19152	19153	19153	19153	19153	19153
Weak id KP Wald F	23.7325	23.7189	23.7170	23.7170	23.7170	23.7170	23.7170
Dep Mean	11.0487	57.7875	0.6680	0.2495	0.0784	0.0081	0.0010
			Panel C: High Severity				
MCR response	0.3225	1.0051	-0.2504	0.1669	0.1449	0.0058	-0.0213
	(0.6875)	(0.7743)	(0.3644)	(0.3330)	(0.2469)	(0.0943)	(0.0455)
Obs	18012	18019	18019	18019	18019	18019	18019
Weak id KP Wald F	22.0178	21.9969	21.9969	21.9969	21.9969	21.9969	21.9969
Dep Mean	12.0577	69.0166	0.6122	0.2645	0.1121	0.0167	0.0037

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A25: Time Spent, Healthcare Access, and Potential Risk of Harm by Count-based Risk of Harm for Mental Health Calls Only

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
			Panel A: Low risk of harm				
MCR response	0.1510	0.8781	0.1186	-0.2287	0.1824*	-0.0489	0.0286**
	(0.6354)	(0.8517)	(0.2091)	(0.1877)	(0.1002)	(0.0495)	(0.0135)
Obs	44897	44913	44913	44913	44913	44913	44913
Weak id KP Wald F	41.2010	41.5160	41.5160	41.5160	41.5160	41.5160	41.5160
Dep Mean	11.4440	38.8582	0.9042	0.0752	0.0149	0.0066	0.0005
			Panel B: Medium risk of harm				
MCR response	0.7495	0.8835	0.1287	-1.1957	0.7836	0.4077	0.2143
	(2.1833)	(2.8304)	(0.8701)	(1.0475)	(0.6025)	(0.3356)	(0.1418)
Obs	27566	27578	27578	27578	27578	27578	27578
Weak id KP Wald F	3.0362	3.0004	3.0004	3.0004	3.0004	3.0004	3.0004
Dep Mean	10.5901	46.9678	0.8589	0.0988	0.0323	0.0120	0.0013
			Panel B: High risk of harm				
MCR response	1.2821	2.1272	-1.0177*	0.3454	0.1821	0.2037	-0.0134
	(1.2530)	(1.5355)	(0.5957)	(0.4540)	(0.3398)	(0.2046)	(0.0785)
Obs	27453	27473	27476	27476	27476	27476	27476
Weak id KP Wald F	8.9046	8.7521	8.7451	8.7451	8.7451	8.7451	8.7451
Dep Mean	11.9551	71.1291	0.7473	0.1551	0.0724	0.0287	0.0050

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A26: Time Spent, Healthcare Access, and Potential Risk of Harm by Count-based Psychiatric Acuity for Mental Health Calls Only

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: Low psychiatric acuity							
MCR response	0.8164	2.7344**	-0.4317	0.1228	0.1052	0.1360	0.0496*
	(0.8003)	(1.1374)	(0.2641)	(0.2222)	(0.1399)	(0.0911)	(0.0292)
Obs	39644	39670	39670	39670	39670	39670	39670
Weak id KP Wald F	22.6416	22.8444	22.8444	22.8444	22.8444	22.8444	22.8444
Dep Mean	11.7986	40.2828	0.8928	0.0819	0.0184	0.0079	0.0005
Panel B: Medium psychiatric acuity							
MCR response	0.9950	-0.2199	0.1402	-0.6097	0.5998**	-0.0517	0.0109
	(1.3418)	(1.7508)	(0.4940)	(0.4610)	(0.2968)	(0.1509)	(0.0509)
Obs	39001	39009	39011	39011	39011	39011	39011
Weak id KP Wald F	10.9793	11.1274	11.1274	11.1274	11.1274	11.1274	11.1274
Dep Mean	10.5896	45.7692	0.8674	0.0933	0.0301	0.0107	0.0012
Panel B: High psychiatric acuity							
MCR response	-0.5347	0.4087	-0.0398	-0.4482	0.2757	0.1132	0.0706
	(0.9900)	(1.1793)	(0.4558)	(0.4272)	(0.2828)	(0.1169)	(0.0513)
Obs	21271	21285	21286	21286	21286	21286	21286
Weak id KP Wald F	12.0031	11.5847	11.5847	11.5847	11.5847	11.5847	11.5847
Dep Mean	11.9696	63.4553	0.7270	0.1808	0.0791	0.0170	0.0034

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. We classify calls with a severity score of 3 or more as high severity, those with severity scores of 1 or 2 as medium severity, and those with a severity score of 0 as low severity. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A27: Time Spent, Healthcare Access, and Potential Risk of Harm by Coding Uncertainty

	(1) Wait Time for Arrive (log min)	(2) Time Spent at Dispatch (log min)	(3) Left at Home (No Transport to Jail, Hospital)	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: Correctly coded calls							
MCR response	0.6408*	1.1061**	-0.3686**	0.2855*	0.0692	0.0329	0.0058
	(0.3808)	(0.4541)	(0.1755)	(0.1592)	(0.1032)	(0.0330)	(0.0120)
Obs	75444	75469	75471	75471	75471	75471	75471
Weak id KP Wald F	80.3391	79.7194	79.7176	79.7176	79.7176	79.7176	79.7176
Patrol officer mean	11.5471	52.5320	0.7307	0.2023	0.0622	0.0079	0.0011
Panel B: Incorrectly coded calls							
MCR response	0.6584	2.2540	0.0566	-0.7770*	0.6283**	0.0346	0.0614
	(1.0440)	(1.4386)	(0.4129)	(0.4108)	(0.2621)	(0.1392)	(0.0517)
Obs	72170	72205	72207	72207	72207	72207	72207
Weak id KP Wald F	16.3092	16.6084	16.6039	16.6039	16.6039	16.6039	16.6039
Patrol officer mean	11.0618	54.6523	0.8271	0.1122	0.0457	0.0172	0.0024

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A28: Characteristics of Compliers In Later Era for Calls Handled by Dispatch Only, not by Right Response

	2022–2024, excluding Right Response Handled Calls	
	Mean (1)	Complier-weighted (2)
Mental health calls		
Low severity	0.218	0.224
	(0.002)	(0.083)
Medium severity	0.260	0.214
	(0.002)	(0.089)
High severity	0.227	0.099
	(0.002)	(0.097)
Suicide-related calls		
Low severity	0.061	0.015
	(0.001)	(0.056)
Medium severity	0.114	0.157
	(0.001)	(0.070)
High severity	0.120	0.291
	(0.001)	(0.083)
Obs	57,340	57,340

Note: Column (1) reports the share of calls in each call type and severity category in the 2022-2024 estimation sample, including only those calls handled by dispatchers, i.e. dropping calls handled by the Right Response Team. Column (2) reports complier-weighted shares for that same sample, i.e., shares weighted by the implicit weights of the 2SLS estimates, following Frandsen *et al.* (2023). Each is the coefficient on MCR response from a 2SLS regression of the category indicator multiplied by MCR response on MCR response, instrumented with recent citywide mental illness dispatches and including the time-varying covariates in Equation 1. Bootstrapped standard errors (500 replications) in parentheses.

Table A29: Involuntary Transport Impact by Call Type, Call Severity and Era

	(1) Low Severity	(2) Mental Health, Medium	(3) Mental Health, High	(4) Suicide-Related, Medium	(5) Suicide-Related, High
			Panel A: 2018-2020		
MCR response	0.0322	0.5128**	-0.1725	-0.0015	0.6888
	(0.1822)	(0.2203)	(0.5331)	(0.3804)	(0.6208)
Obs	18262	14940	9626	8545	6288
Weak id KP Wald F	12.9542	17.3476	4.3482	8.3288	4.5294
Patrol mean	0.0283	0.0317	0.0621	0.0730	0.1041
			Panel B: 2022-2024		
MCR response	0.5912*	0.3908	0.6136	0.0815	0.1171
	(0.3446)	(0.3463)	(0.6438)	(0.3348)	(0.2759)
Obs	17734	16415	15845	7420	9239
Weak id KP Wald F	6.5011	6.0585	2.7658	10.3087	15.7912
Patrol mean	0.0277	0.0359	0.0734	0.0862	0.1201

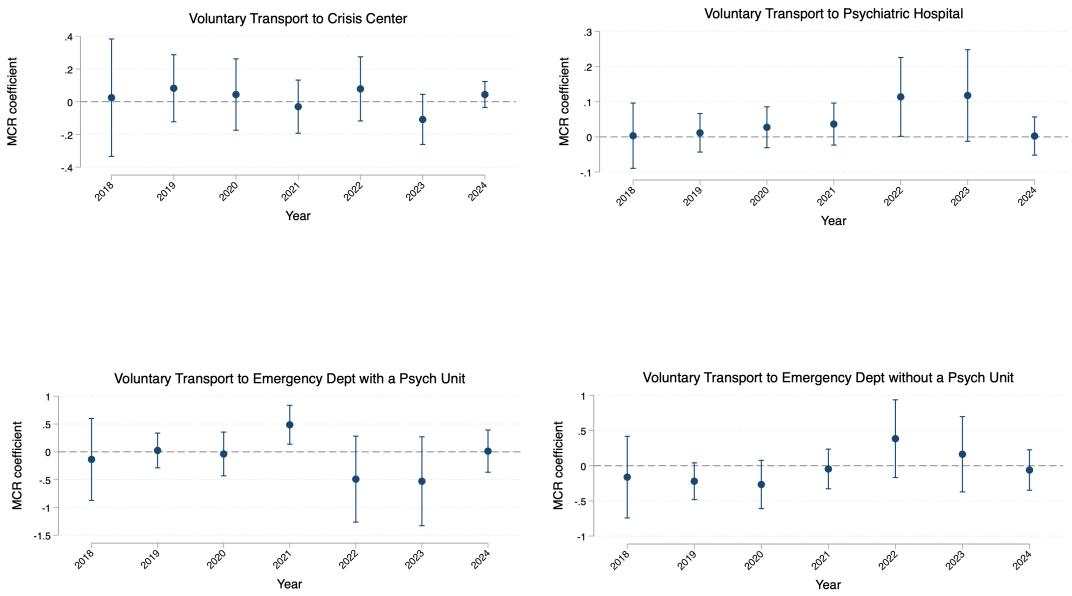
Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A30: Voluntary Transport Location Impact by Era

	Voluntary Transport to:			
	(1) Crisis Center	(2) Psych Hospital	(3) Emergency Dept with No Psych Unit	(4) Emergency Dept with Psych Unit
Panel A: 2018-2020				
MCR response	0.0453	0.0167	-0.2476*	-0.0242
	(0.0920)	(0.0244)	(0.1338)	(0.1588)
Obs	57712	57712	57712	57712
Weak id KP Wald F	47.2609	47.2609	47.2609	47.2609
Dep Mean	0.0205	0.0017	0.0592	0.0844
Panel B: 2021 only				
MCR response	-0.0303	0.0364	-0.0462	0.4867**
	(0.0988)	(0.0363)	(0.1713)	(0.2129)
Obs	23160	23160	23160	23160
Weak id KP Wald F	29.0908	29.0908	29.0908	29.0908
Dep Mean	0.0163	0.0026	0.0610	0.0808
Panel C: 2022-2024				
MCR response	0.0363	0.0625**	0.1221	-0.2491
	(0.0430)	(0.0282)	(0.1387)	(0.1924)
Obs	66806	66806	66806	66806
Weak id KP Wald F	36.5404	36.5404	36.5404	36.5404
Dep Mean	0.0040	0.0022	0.0548	0.0879

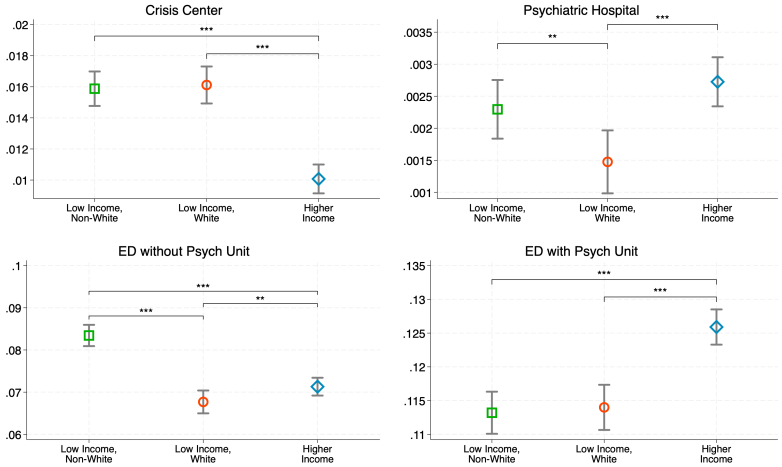
Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A8: Impact on Voluntary Transport Location by Year



Note: The coefficient listed is β from Equation 1, obtained through year-specific two-stage least squares estimation. 90% confidence intervals are shown, constructed via robust standard errors.

Figure A9: Pooled Voluntary and Involuntary Transport Location by Neighborhood Category



Note: In all subfigures, neighborhood-specific means have been calculated in the universe of all mental illness dispatches by standard police (i.e., excluding dispatches with MCR response). Stars indicate mean difference t-test p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. T-tests of time means were done with log time.

Table A31: Transport Location Impact by Distance to Facility

	Voluntary and Involuntary Transports to:	
	(1) Crisis Center	(2) Emergency Dept with Psych Unit
Panel A: Crises close to the relevant facilty		
MCR response	-0.0079	0.0650
	(0.0949)	(0.1788)
Obs	73838	73843
Weak id KP Wald F	44.7586	58.6085
Patrol Officer Mean	0.0194	0.1354
Panel B: Crises far from relevant facility		
MCR response	0.0497	0.3487
	(0.0579)	(0.2127)
Obs	73837	73832
Weak id KP Wald F	45.0049	32.3874
Patrol Officer Mean	0.0073	0.1023

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Panel A includes crises that are within 6.09 miles of the free Crisis Center (the median distance in our analysis sample) in Column 1, and crises that are within 7.28 miles of an emergency department with a psychiatric unit (the median distance in our analysis sample) in Column 2. Panel A includes crises that are further from the free Crisis Center (Column 1) or an an emergency department with a psychiatric unit (Column 2). Outcomes indicate transport location for both voluntary and involuntary transports. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A32: Time Spent, Healthcare Access, and Potential Risk of Harm Impact (IV) by Length of Department Tenure

	(1) Time to arrive (log min)	(2) Time at dispatch (log min)	(3) No transport to jail or hospital	(4) Voluntary Transport	(5) Involuntary Transport	(6) Jailed	(7) Force Used
Panel A: MCR Officer Tenure in Department ≤ 19							
MCR response	0.9231	1.6351**	-0.1024	-0.2075	0.2840*	0.0251	0.0351
	(0.6219)	(0.7971)	(0.2646)	(0.2355)	(0.1501)	(0.0699)	(0.0247)
Obs	142219	142275	142279	142279	142279	142279	142279
Weak id KP Wald F	65.5349	65.3517	65.3560	65.3560	65.3560	65.3560	65.3560
Patrol officer mean	11.3053	53.5887	0.7788	0.1574	0.0540	0.0125	0.0018
MCR program tenure	1.5447	1.5458	1.5458	1.5458	1.5458	1.5458	1.5458
Panel B: MCR Officer Tenure in Department > 19							
MCR response	1.2428	2.4480**	-0.1542	-0.2122	0.3009	0.0597	0.0532
	(0.9346)	(1.2118)	(0.3933)	(0.3509)	(0.2236)	(0.1039)	(0.0378)
Obs	141783	141838	141842	141842	141842	141842	141842
Weak id KP Wald F	30.0619	30.1278	30.1198	30.1198	30.1198	30.1198	30.1198
Patrol officer mean	11.3053	53.5887	0.7788	0.1574	0.0540	0.0125	0.0018
MCR program tenure	2.6362	2.6373	2.6373	2.6373	2.6373	2.6373	2.6373

Note: The coefficient listed is β from Equation 1, obtained through two-stage least squares estimation. Tenure in the department is calculated as the difference between the date when the senior-most MCR officer was hired by CPD and the date of the incident. In 87 percent of MCR responses, only 1 MCR team responds, making the "senior-most MCR officer" responding the only MCR officer responding. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A33: Effect of MCR Response on Resource Utilization for Officers Only

	(1) # Officers	(2) # Patrol Officers	(3) # Unrequired Officers
MCR response	0.6947	-0.0608	0.3915
	(0.7476)	(0.7788)	(0.7110)
Obs	144056	144056	143612
Weak id KP Wald F	87.4676	87.4676	87.6268
Patrol officer mean	2.6334		0.6075

Note: These estimates are obtained after controlling for police code categories originally assigned, as well as day-of-week × dispatch hour, month, and year fixed effects. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. We report the robust KP Wald F Kleibergen-Paap Wald rk F statistic, which is same as the effective F-statistic of Olea and Pflueger (2013) in just identified settings, and can be compared with the Stock and Yogo (2005) critical values. The Stock-Yogo weak ID test critical values for 10%, 15%, 20%, and 25% maximal IV size is 16.38, 8.96, 6.66, and 5.53 respectively.

Table A34: Future Calls from the Neighborhood of Call i

	Any Call Within 30 days			Any Call Within 180 days		
	(1) 50m	(2) 100m	(3) 200m	(4) 50m	(5) 100m	(6) 200m
MCR response	0.2293	0.0739	0.0515	0.2060**	0.0300	-0.0088
	(0.1822)	(0.1253)	(0.0578)	(0.0993)	(0.0422)	(0.0121)
Obs	146,059	146,059	146,059	137,491	137,491	137,491
Weak id KP Wald F	88.5078	88.5078	88.5078	81.6081	81.6081	81.6081
Officer mean	0.7207	0.8935	0.9797	0.9443	0.9904	0.9991

Note: These estimates are obtained after controlling for police code categories originally assigned, as well as day-of-week × dispatch hour, month, and year fixed effects. Robust Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Healthcare and local support organizations includes calll from healthcare facilities, homeless shelters, rehabilitation facilities, libraries, and assisted living facilities. Commercial and educational institutions includes calls from local businesses, groceries, restaurants, public spaces, and schools.

B Appendix B: Data Sources and Description

Behavioral Health Classification

Columbus police use five primary police codes for behavioral health calls, but these codes do not capture all such incidents. Call-takers typically assign one of the following codes to behavioral health calls: 10-16B ("disturbance/mental"), 10-47A ("suicide attempt"), 10-19 ("intoxicated person"), 10-34O ("overdose, accidental"), or 10-44B ("indecent exposure"). However, when an incident overlaps multiple call categories, internal dispatch protocols lead call-takers to assign a different code depending on the nature of the call, its severity, or the involvement of weapons. For example, a call from a mother reporting that her daughter is cutting her wrist may be coded as a priority one "cutting/stabbing" rather than a priority two "suicide attempt" or "disturbance/mental." Similarly, a call about a person in a psychotic episode who is behaving combatively toward family members may be coded as "domestic violence" rather than "disturbance/mental." Code assignment also varies across operators, depending on their training, experience, and time of shift. Relying solely on dispatcher-assigned codes would therefore omit a substantial share of behavioral health calls from our analysis.

To address this limitation, we identify behavioral health calls in two steps, combining the assigned police codes with the call remarks recorded by the operator. First, we classify as behavioral health all calls carrying one of the five codes above. Second, we review the remarks on these calls and compile a lexicon of keywords and phrases (e.g., "high on meth," "bipolar") and paired terms (e.g., "passed out" and "needle") that indicate behavioral health crises. We then search for these terms in the remarks field across the entire dataset and classify matching calls into one of four categories: substance use, suicide, overdose, or mental health. We refine the lexicon iteratively until a satisfactory share of calls initially coded 10-16B, 10-44B, 10-47A, 10-34O, or 10-19 is also identified as behavioral health based on remarks alone.

Table B1: Categorical Designation of Calls by Dispatcher-Assigned Police Code

Remarks based	Tencode based				
	Intoxicated Person	Disturb Mental	Overdose Accidental	Suicide Attempt	Other Tencodes
Substance Use	3,534	1,598	1,054	1,202	101,688
Mental Illness	66	30,828	123	2,624	46,685
Overdose	8	72	14,374	212	3,628
Suicide	11	7,803	207	24,613	9,346
All mental health calls	3,619	40,301	15,758	28,651	161,347
N	4,440	58,639	20,055	34,505	2,127,117

NOTE: This table shows behavioral health call classifications based on both the assigned police code and call remarks, for nearly 2.8 million calls successfully dispatched within Columbus police precincts between June 2018 and December 2024. Column 2 reports classifications based on call remarks for calls originally coded as either "disturbance - mental" (10-16B) or "indecent exposure" (10-44B) by dispatch employees.

In the final classification, we designate a call as behavioral health if either its dispatch code or its remarks identify it as such. This yields 249,676 behavioral health calls, of which

73,349 receive the same classification under both approaches (Table B1 details the overlap between the two). Where the two methods assign different categories, we give precedence to the remarks-based category, since it better captures the context of the incident. For example, we treat a call that the call-taker assigns 10-16B ("disturbance/mental") but is classified as suicide based on call remarks as a suicide related call in our analysis.

Use of force data

The Columbus Division of Police recognizes nine levels of force, ranging from 0 to 8. Level 0 includes officer presence, verbal and nonverbal commands, searches, handcuffing, displaying or sparking a Taser for compliance, displaying a firearm, use of flashbangs and multiple baton rounds as diversions, and the use of a Long Range Acoustic Device warning tone; Level 1 involves empty-hand control, pressure points, grounding techniques, joint manipulations, and pushes with objects such as bicycles, riot shields, or batons; Level 2 includes the use of chemical spray; Level 3 involves the use of electronic devices (e.g., electronic custody belts or Conducted Energy Weapons such as Tasers) or air launchers; Level 4 involves use of hard empty-hand control (e.g., strikes, punches, or kicks); Level 5 includes use of impact weapons (e.g., batons or flashlights); Level 6 involves police K-9 (dog) bites; Level 7 includes use of less-lethal weapons (e.g., beanbag rounds or multiple baton rounds); and finally level 8 involves the use of deadly force (Columbus Police Division, 2023).

Records covering all use-of-force incidents from January 2015 to December 2024 were obtained directly from the Columbus Police Department (CPD) through a public records request. The use of force dataset is created and maintained by the Internal Affairs division of CPD. The data provided by Internal Affairs division did not have details on low and high levels of uses of force i.e. level 0, 1 and 8. It mainly documents level 2-7 and only select Level 0 and 1 incidents where an injury is recorded. Thus, all use of force in our dataset can be thought of as "serious", producing harm or at least a serious risk of harm to the individual receiving force. Each record is at the incident–officer–suspect level, meaning it captures all instances of force used by a particular officer against a person for a given incident (identified by a unique event number). We distill these data down to a binary indicator of any ("serious") use of force being used, against anyone by any officer, during a given dispatch.

We further suplpemented this data with use of force data recorded by officers via paper form, and data from those paper forms are input into an excel database by police Internal Affairs. Up until 2023, the Internal affirs was not required to report low levels of use of force but from 2023 onwards as a result of Department of Justice Review, level 0 and 1 incidents are now also to be reported. It is possible that officers do not always record force incidents but there are penalties for failing to do so, and so we do not expect there to be a major form of measurement error. However, digitization work undertaken by Ohio State University investigators has found that a very large proportion of traffic stops reported by police on paper forms do not make it into the Internal Affairs data on traffic stops, perhaps because the data entry process is so onerous. Even storage of these paper forms presents a serious problem. These same investigators have obtained and encoded all paper forms for use of force in 2023 and 2024 allowing us to compare the universe of paper-recorded use of force incidents with the subset of incidents that are in

the 2023 and 2024 Internal Affairs datasets.

Determining final transport variables

We construct outcomes capturing police transports to healthcare facilities and to jail from the transport field in the CAD dispatch data. This field records every change in transport location in chronological order, as entered by the responding officers after arriving on scene (the last column of Table A1 illustrates how these entries appear). The field is blank when no transport occurs.

Police transports are not limited to healthcare facilities or jail. Depending on the circumstances, officers also take individuals to homeless shelters, Franklin County Children's Services, the property room for evidence storage, police substations for paperwork, headquarters for identification, or another address or residence requested by someone involved in the incident.

To isolate the relevant destinations, we compile a list of healthcare and jail locations in Columbus using input from the Columbus Division of Police and supplemental online searches. Crisis centers comprise Netcare and RI International. Private psychiatric hospitals include Twin Valley Behavioral Healthcare Hospital, Columbus Springs Dublin, Columbus Springs East, SUN Behavioral Health, SEQUEL-Pomegranate Health Systems, RiverVista Behavioral Hospital, and Mount Carmel Behavioral Health. Emergency departments with a psychiatric unit include Riverside Methodist Hospital, Ohio State University Wexner Medical Center, Nationwide Children's Hospital, and Mount Carmel West Hospital, which had a psychiatric unit until 2019. Emergency departments without a psychiatric unit include Dublin Methodist Hospital, Grant Medical Center, Doctors Hospital, Mount Carmel, and other OhioHealth and Ohio State University hospital campuses. Transports to jail are identified from text references in the transport field such as "jail," "Jackson," "Pike," or "county."

We match this list to the text in the transport field and generate indicators for transports to crisis centers, private psychiatric hospitals, emergency departments (with and without psychiatric units), and jail, using the sequence of recorded locations to identify the final destination for each call. The sequence matters because a facility mention in the transport field does not by itself indicate the final destination. First, the jail does not accept individuals who need medical attention until a hospital has cleared them, so a hospital visit may precede a jail booking within the same incident. Second, the intended destination often differs from the actual one: an officer may plan to take a person in crisis to a free crisis center such as Netcare, but if the facility is at capacity, the person is taken to a public hospital instead. Where the transport field is blank but the call remarks indicate a drop-off, we use the call remarks to identify the destination.

C Appendix C: Complier-Weighted Characteristics and Reweighting

Complier-weighted shares with a single instrument and controls

Frandsen *et al.* (2023) show that, with judge indicators as instruments, the coefficient on treatment D_i from a 2SLS regression of $X_i D_i$ on D_i recovers $E[\omega_i X_i]/E[\omega_i]$, the average of a pre-determined characteristic X_i weighted by each individual's implicit weight ω_i in the 2SLS estimand. Their derivation abstracts from covariates. We adapt it to our setting, with a single count instrument Z_i (recent citywide mental illness and suicide dispatches), MCR response D_i, and the time-varying controls C_i in Equation 1. Let $D_i(z)$ denote call i's potential MCR response at instrument value z, and let $\tilde{Z}_i = Z_i - L[Z_i \mid C_i]$ denote the instrument residualized on the controls. We assume (i) $Z_i \perp \big(X_i, D_i(\cdot)\big) \mid C_i$; (ii) monotonicity, so that $D_i(z)$ is weakly decreasing in z for all i; and (iii) $E[Z_i \mid C_i]$ is linear in C_i (Blandhol *et al.*, 2026).

By Frisch–Waugh–Lovell, the 2SLS coefficient from a regression of $W_i = X_i D_i$ on D_i, controlling for C_i, is

$$\beta = \frac{E[\tilde{Z}_i X_i D_i]}{E[\tilde{Z}_i D_i]} = \frac{E\big[X_i\, E[\tilde{Z}_i D_i(Z_i) \mid X_i, D_i(\cdot), C_i]\big]}{E\big[E[\tilde{Z}_i D_i(Z_i) \mid X_i, D_i(\cdot), C_i]\big]} = \frac{E[\omega_i X_i]}{E[\omega_i]},$$

where, by (i) and (iii),

$$\omega_i = \mathrm{Cov}\big(Z_i, D_i(Z_i) \mid C_i\big).$$

That is, ω_i is the covariance, within call i's control cell, between the instrument and call i's potential MCR response. Under (ii), $\omega_i \le 0$ for all i, so β is a proper weighted average of X_i; the sign cancels in the ratio. Calls whose MCR response does not vary with the instrument have $\omega_i = 0$. If the first stage is linear in Z_i within control cells, ω_i is proportional to the Frandsen *et al.* (2023) weight, with the constant of proportionality cancelling in the ratio. When X_i is an indicator for call cell c, β is the complier share of cell c, $s_c = E[\omega_i 1\{c_i = c\}]/E[\omega_i]$, and the shares sum to one across mutually exclusive cells. We estimate shares separately within each era.

Reweighting

Let $s_{c,1}$ and $s_{c,3}$ denote the complier share of cell c in 2018–2020 and 2022–2024. Panel D of Table 9 weights each 2022–2024 call in cell c by $k_c = s_{c,1}/s_{c,3}$. Holding the residualized instrument fixed, the reweighted share of cell c is

$$\frac{k_c\, s_{c,3}}{\sum_{c'} k_{c'}\, s_{c',3}} = \frac{s_{c,1}}{\sum_{c'} s_{c',1}} = s_{c,1}.$$

In practice, reweighting also slightly changes the residualization of the instrument on the controls, so the match is approximate. Re-estimating the 2022–2024 complier shares in the reweighted sample yields shares within 0.002 of the 2018–2020 targets for every cell.

Inference

Standard errors in Table 10 are bootstrapped with 500 replications, resampling calls within era, following Dobbie *et al.* (2018) and Dahl *et al.* (2014). Standard errors in Panel D of Table 9 treat the weights k_c as fixed and therefore understate uncertainty.

References

Adger, C., Ross, M. and Sloan, C. (2022) The effect of field training officers on police use of force, *United States Military Academy at West Point Working Paper*.

Akther, S. F., Molyneaux, E., Stuart, R., Johnson, S., Simpson, A. and Oram, S. (2019) Patients' experiences of assessment and detention under mental health legislation: systematic review and qualitative meta-synthesis, *BJPsych Open*, **5**, e37.

Alcohol, Drug and Mental Health (2024) ADAMH Franklin County 2024 Community Needs Assessment.

Angrist, J. and Krueger, A. B. (1994) Why do world war ii veterans earn more than nonveterans, *Journal of Labor Economics*, **12**.

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

Ba, B. A., Chen, P., Cheng, T., Eies, M. C. and Holz, J. E. (2025) What is the best response? examining the impact of police and their alternatives.

Bhuller, M., Dahl, G. B., Løken, K. V. and Mogstad, M. (2020) Incarceration, recidivism, and employment, *Journal of political economy*, **128**, 1269–1324.

Blandhol, C., Bonney, J., Mogstad, M. and Torgovitsky, A. (2026) When is tsls actually late?, *Review of Economic Studies*.

Boscarato, K., Lee, S., Kroschel, J., Hollander, Y., Brennan, A. and Warren, N. (2014) Consumer experience of formal crisis-response services and preferred methods of crisis intervention, *International journal of mental health nursing*, **23**, 287–295.

Boudreaux, E. D., Brown, G. K., Stanley, B. *et al.* (2017) Computer administered safety planning for individuals at risk for suicide: Development and usability testing, *Journal of Medical Internet Research*, **19**, e149.

Bove, V. and Gavrilova, E. (2017) Police officer on the frontline or a soldier? the effect of police militarization on crime, *American Economic Journal: Economic Policy*, **9**, 1–18.

Brody, D. J. and Hughes, J. P. (2025) Depression prevalence in adolescents and adults: United states, august 2021–august 2023, NCHS Data Brief 527, National Center for Health Statistics (U.S.), Hyattsville, MD.

Centers for Disease Control and Prevention (2024) Youth risk behavior survey data summary & trends report: 2013–2023, Tech. rep., U.S. Department of Health and Human Services, Atlanta, GA.

Centers for Medicare & Medicaid Services (2021) Medicaid guidance on the scope of and payments for qualifying community-based mobile crisis intervention services, State Health Official Letter #21-008, december 28, 2021.

Chan, D. C. (2018) The efficiency of slacking off: Evidence from the emergency department, *Econometrica*, **86**, 997–1030.

Cho, S., Gonçalves, F. and Weisburst, E. (2024) The impact of fear on police behavior and public safety, *Review of Economics and Statistics*, pp. 1–45.

Chyn, E., Frandsen, B. and Leslie, E. (2025) Examiner and judge designs in economics: A practitioner's guide, *Journal of Economic Literature*, **63**, 401–439.

Columbus Police Division (2023) Columbus Police Division Directives: Use of Force, https://www.columbus.gov/files/sharedassets/city/v/2/public-safety/police/directives/divisiondirective2.01.pdf [Accessed: October 19 , 2024].

Dahl, G. B., Kostøl, A. R. and Mogstad, M. (2014) Family welfare cultures, *The Quarterly Journal of Economics*, **129**, 1711–1752.

Davis, J., Norris, S., Schmitt, J., Shem-Tov, Y. and Strickland, C. (2025) Mobile crisis response teams support better policing: Evidence from cahoots.

Dee, T. S. and Pyne, J. (2026) Emergency mental health co-responders reduce involuntary psychiatric detentions in the usa, *Nature Human Behaviour*, **10**, 148–155.

Dobbie, W., Goldin, J. and Yang, C. S. (2018) The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges, *American Economic Review*, **108**, 201–240.

Emanuel, N., Welle, P. and Bolotnyy, V. (2025) A danger to self and others: Health and criminal consequences of involuntary hospitalization, *FRB of New York Staff Report*.

Ettman, C. K., Cohen, G. H., Abdalla, S. M. *et al.* (2022) Persistent depressive symptoms during COVID-19: A national, population-representative, longitudinal study of U.S. adults, *The Lancet Regional Health – Americas*, **5**, 100091.

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

Frandsen, B., Lefgren, L. and Leslie, E. (2023) Judging judge fixed effects, *American Economic Review*, **113**, 253–277.

Freedman, S., Golberstein, E., Huang, T.-Y., Satin, D. J. and Smith, L. B. (2021) Docs with their eyes on the clock? The effect of time pressures on primary care productivity, *Journal of Health Economics*, **77**, 102442.

Hefner, J. L., Noghrehchi, P., David, S., Pandey, S. and Bevis, L. E. (2026) A community-engaged qualitative study of police response to 911 calls for behavioral health crises, *Progress in Community Health Partnerships: Research, Education, and Action*, **20**, 13–22.

Huttunen, K., Kaila, M., Macdonald, D. C. and Nix, E. (2026) Punishing financial crimes: The impact of prison sentences on defendants and their colleagues, *American Economic Journal: Economic Policy*, **18**, 477–510.

Jones, N., Gius, B., Shields, M., Florence, A., Collings, S., Green, K., Watson, A. and Munson, M. (2022) Youths' and young adults' experiences of police involvement during initiation of involuntary psychiatric holds and transport, *Psychiatric services*, **73**, 910–917.

Jones, N., Gius, B. K., Shields, M., Collings, S., Rosen, C. and Munson, M. (2021) Investigating the impact of involuntary psychiatric hospitalization on youth and young adult trust and help-seeking in pathways to care, *Social Psychiatry and Psychiatric Epidemiology*, **56**, 2017–2027.

Kerwin, J., Rostom, N. and Sterck, O. (2026) Striking the right balance: Why standard balance tests over-reject the null, and how to fix i, *Working Paper*.

Lowder, E. M., Grommon, E., Bailey, K. and Ray, B. (2024) Police-mental health co-response versus police-as-usual response to behavioral health emergencies: A pragmatic randomized effectiveness trial, *Social Science & Medicine*, **345**, 116723.

Marcus, N. and Stergiopoulos, V. (2022) Re-examining mental health crisis intervention: A rapid review comparing outcomes across police, co-responder and non-police models, *Health & social care in the community*, **30**, 1665–1679.

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

Miller, A. R. and Segal, C. (2019) Do female officers improve law enforcement quality? effects on crime reporting and domestic violence, *The review of economic studies*, **86**, 2220–2247.

Mission Critical Partners (2025) Alternative Response Assessment Final Report.

Namigadde, A. (2019) When mental health crises arise, Columbus police brings social workers along, WOSU 89.7 NPR News, accessed October 25, 2026.

Neprash, H. T., Mulcahy, J. F., Cross, D. A., Gaugler, J. E., Golberstein, E. and Ganguli, I. (2023) Association of primary care visit length with potentially inappropriate prescribing, *JAMA Health Forum*, **4**, e230052.

Norris, S., Pecenco, M. and Weaver, J. (2021) The effects of parental and sibling incarceration, *The American Economic Review*, **111**, 2926–2963.

Olea, J. L. M. and Pflueger, C. (2013) A robust test for weak instruments, *Journal of Business & Economic Statistics*, **31**, 358–369.

Pope, L. G., Patel, A., Fu, E., Zingman, M., Warnock, A., Ellis, S., Ashekun, O., Watson, A., Wood, J. and Compton, M. T. (2023) Crisis response model preferences of mental health care clients with prior misdemeanor arrests and of their family and friends, *Psychiatric Services*, **74**, 1163–1170.

Richmond, J. S., Berlin, J. S., Fishkind, A. B., Holloman, G. H., Zeller, S. L., Wilson, M. P., Rifai, M. A. and Ng, A. T. (2012) Verbal de-escalation of the agitated patient: Consensus statement of the American Association for Emergency Psychiatry Project BETA de-escalation workgroup, *Western Journal of Emergency Medicine*, **13**, 17–25.

Rivera, R. (2025) Do peers matter in the police academy?, *American Economic Journal: Applied Economics*, **17**, 127–164.

Roman, M., Fredriksson, K., Cassella, C., Epp, D. A. and Walker, H. L. (2023) The george floyd effect: How protests and public scrutiny changed police behavior, *Perspectives on Politics*, pp. 1–21.

Rozema, K. and Schanzenbach, M. (2019) Good cop, bad cop: Using civilian allegations to predict police misconduct, *American Economic Journal: Economic Policy*, **11**, 225–268.

Schwartz, J. C. (2014) Police indemnification, *NYUL Rev.*, **89**, 885.

Shapiro, G. K., Cusi, A., Kirst, M., O'Campo, P., Nakhost, A. and Stergiopoulos, V. (2015) Co-responding police-mental health programs: A review, *Administration and policy in mental health and mental health services research*, **42**, 606–620.

Silver, D. (2021) Haste or waste? Peer pressure and productivity in the emergency department, *Review of Economic Studies*, **88**, 1385–1417.

Stanley, B. and Brown, G. K. (2012) Safety planning intervention: A brief intervention to mitigate suicide risk, *Cognitive and Behavioral Practice*, **19**, 256–264.

Stock, J. and Yogo, M. (2005) Asymptotic distributions of instrumental variables statistics with many instruments, *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, **6**, 109–120.

Sun, Y., Wu, Y., Fan, S. *et al.* (2023) Comparison of mental health symptoms before and during the COVID-19 pandemic: Evidence from a systematic review and meta-analysis of 134 cohorts, *BMJ*, **380**, e074224.

Uding, C. V., Moon, H. R. and Lum, C. (2025) The status of co-responders in law enforcement: findings from a national survey of law enforcement agencies, *Policing: An International Journal*, **48**, 69–97.

Van Kippersluis, H. and Rietveld, C. A. (2018) Beyond plausibly exogenous, *The Econometrics Journal*, **21**, 316–331.

Watson, A. C., Compton, M. T. and Pope, L. G. (2019) Crisis response services for people with mental illnesses or intellectual and developmental disabilities: A review of the literature on police-based and other first response models, *Vera Institute of Justice*.

Weisburd, S. (2021) Police presence, rapid response rates, and crime prevention, *Review of Economics and Statistics*, **103**, 280–293.

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

Williams Jr, M. C., Weil, N., Rasich, E. A., Ludwig, J., Chang, H. and Egrari, S. (2021) Body-worn cameras in policing: Benefits and costs.

Yang, S.-M., Gill, C. E., Lu, Y.-F., Azam, M. and Kanewske, L. C. (2024) A police-clinician co-response team to people with mental illness in a suburban-rural community: A randomized controlled trial, *Journal of Experimental Criminology*, pp. 1–22.

Yang, S.-M. and Lu, Y.-F. (2024) Evaluating the effects of co-response teams in reducing subsequent hospitalization: A place-based randomized controlled trial, *Policing: A Journal of Policy and Practice*, **18**, paad080.