Stanford | John W. Gardner Center for Youth and Their Communities

REPORTS IN THIS SERIES

Impact Report

Implementation Reports
Dispatch
Co-response
Follow-up & Continuity of Care

Background Reports
Theory of Change
Program Impacts: Technical Report

PROGRAM IMPACTS

Technical Report

October 2024

San Mateo County, California
Community Wellness and Crisis Response Team

Pilot Program: December 2021 through June 2024

Thomas S. Dee, Stanford University

Barnett Family Professor
Graduate School of Education

Faculty Director
John W. Gardner Center for Youth and Their Communities

Jaymes Pyne, Stanford University

Senior Research Associate
John W. Gardner Center for Youth and Their Communities

Create knowledge.
Ignite change.

The Community Wellness and Crisis Response Team (CWCRT) pilot program—operating in the four largest cities in San Mateo County, California—provides a mental health clinician co-responding with a sworn law enforcement officer to 911 calls for service involving someone experiencing a mental health-related crisis.

San Mateo County has used several models to address community mental health-related crises for more than two decades. These programs include law enforcement first responders who undergo crisis intervention training (CIT), Psychiatric Emergency Response Teams (PERT), and the San Mateo Assessment and Referral Team (SMART). Each of these programs provides support for community members in crisis situations and each relies on partnerships across multiple agencies.

A combination of factors, including input from community organizations and constituents, has led county leaders to seek even more approaches to address mental health crises. To do so, the San Mateo Executive's Office collaborated with the county's Behavioral Health and Recovery Services (BHRS), StarVista (a nonprofit offering counseling and crisis prevention services), and police agencies within the county's four largest cities: Daly City, Redwood City, San Mateo, and South San Francisco. In December 2021, this partnership began implementing the Community Wellness and Crisis Response Team (CWCRT) Pilot Program, which employs a co-responder model pairing sworn law enforcement officers with mental health clinicians in a first-responder framework.

The county's two-year pilot of the CWCRT program embeds a StarVista mental health clinician within each of the four police departments. The program dispatches each clinician in the assigned city to calls with a known or suspected mental health component—at the same time that a police officer is dispatched—and the clinician arrives on scene in a separate, unmarked vehicle and in plain clothes once the scene has been deemed safe by officers.

A key goal of this collaborative effort is to combine the expertise and resources of both law enforcement and mental health professionals to best serve the public in a timely manner and to improve the outcomes of those served. Dispatchers for 911 calls or law enforcement officers responding to 911 calls request that CWCRT mental health clinicians go out to eligible incidents in each of the four cities during the program's hours of operation (typically Monday to Friday, 9 a.m. to 6 p.m.). Over a 24-month period, police and dispatch records indicate that CWCRT clinicians co-responded with police officers to over 1,500 qualified emergency calls for service in the four participating cities.

Current Study

Our study of the CWCRT program's effects relies on two overarching research agendas. The first research agenda seeks to understand the CWCRT program's total effects, which include both combined programmatic effects at the moment of emergency response (i.e., the incident-level effects) and broad changes in the communities served over time (i.e., community-level effects).

The community-level effects research agenda relies on police-area-by-month panel data on outcomes (e.g., program-related calls for service, police cases, criminal offenses, and involuntary psychiatric detentions) to estimate the impact of the pilot program. This "difference-in-differences" (DiD) approach effectively compares the change in outcomes among treated police areas to the contemporaneous change among untreated police areas. A key assumption of DiD models is that untreated potential outcomes between the two groups proceed in parallel before exposure to the policy or program shock. We identify four domains of study: frequency of program-related incidents/calls for service received by emergency dispatchers; criminal offenses related to the program's stated goals; arrests related to the program's stated goals; and involuntary psychiatric detentions. Our focal timeframe and unit of analysis is program-related calls for service per police area in 48 distinct police areas (26 treatment, 22 comparison) within nine cities in San Mateo County, California, in a given month from January 1, 2019, through June 30, 2023. Our DiD design strategy allows us to take advantage of our panel dataset in months surrounding implementation. This time period represents the first year and a half of the pilot phase of CWCRT up to the last round of full data collection (December 2021–June 2023) and the three years prior to the pilot launch (January 2019–November 2021).

We propose four hypotheses related to our preregistered outcome domains for the total effects of the CWCRT program. First, we anticipate that the CWCRT program will reduce the prevalence of program-related emergency calls for service, particularly as those services persist over time. To the extent that calls for service descriptions provide adequately strong signals for CWCRT activities, we anticipate the frequency of certain call types will fall modestly as community members receive adequate, regular care for their mental health needs. Second, we expect to see decreases in the frequency of certain types of criminal offenses that are most related to CWCRT response priorities. Since the CWCRT Theory of Change (Gardner Center, 2024d) proposes that mental health clinicians who co-respond with police officers effectively advocate for routing individuals to mental health services and away from the criminal justice system, we anticipate fewer responses ending in a criminal case being opened when a clinician arrives with police officers. Third, we similarly expect that because criminal offense frequencies will be reduced, the number of arrests associated with CWCRT responses will go down as well. Those effects may be more than additive, as well; even if a criminal case is opened, we anticipate that the presence of a mental health clinician may lead to less-severe consequences for individuals in crisis (i.e., experiencing a warning, citation, or no action taken rather than arrest during the course of the emergency response).

Finally, we expect a decrease in the frequency of involuntary psychiatric detentions in treated police areas due to the CWCRT program. Involuntary psychiatric detentions are emergency holds in which the individual subjects of a first response are directed to psychiatric services rather than the criminal justice system (e.g., arrest, temporary detention) to receive more appropriate mental health care. The unique skill sets of CWCRT clinicians may mean that they more readily identify the nuances of situations in which an individual may be safe to release in their community (perhaps under the care of a family member) rather than being mandated to psychiatric services. This nuanced approach to mental health care may free up emergency psychiatric services for others and also keep low-threat individuals from the negative consequences of such holds (i.e., coerced care, loss of autonomy). Additionally, if assisted individuals

receive the ongoing help they need by coming into contact with the CWCRT program (including follow-up by program clinicians), they may be less likely to find themselves in mental health crises necessitating involuntary psychiatric detentions in future months. We state these four hypotheses as follows:

Hypothesis A1
Police areas participating in the CWCRT Pilot Program will experience reduced frequency of program-related calls for service.

Hypothesis A2
Police areas participating in the CWCRT Pilot Program will experience reduced frequency of criminal offenses (i.e., the opening of a criminal case in response to a documented complaint, warning, citation, or arrest), related to CWCRT's programmatic focus (i.e., individuals in mental health distress).

Hypothesis A3
Police areas participating in the CWCRT Pilot Program will experience reduced frequency of arrests, related to CWCRT's programmatic focus (i.e., individuals in mental health distress).

Hypothesis A4
Police areas participating in the CWCRT Pilot Program will experience reduced frequency of individuals undergoing an involuntary psychiatric detention.

Our second research agenda focuses only on data from three of the four treatment cities during the months when the CWCRT program is active—identifying the immediate effect of a co-response team on proximate, incident-level outcomes (e.g., involuntary psychiatric detentions, police cases, criminal offenses, arrests).[1] This approach relies on a fixed-effects design applied to an analytical sample of incidents in which a dispatcher or police officer indicated a co-responder was relevant. Conditioning on fixed effects unique to location, day of the week, hour of the day, and year/month, this approach compares outcomes among eligible incidents in which a co-response team showed up with eligible incidents not receiving a co-response (e.g., due to capacity constraints).

This research agenda relies on incident-level data for the number of criminal offenses, number of arrests, and number of involuntary psychiatric detentions for all reported incidents deemed "CWCRT eligible" by a dispatcher or police officer, regardless of whether a clinician was available to respond to the incident. Our focal timeframe and unit of analysis is CWCRT-eligible incidents/calls (determined by a dispatcher or officer) for service in three of the four cities participating in the CWCRT program from December 15, 2021, through December 31, 2023. Our fixed-effects design strategy effectively compares outcomes conditional on effects associated with police area, year, month of year, day of week, and/or hour of day. The remaining participating city we omitted from the sample only recorded whether a clinician responded to an incident and did not flag incidents that were considered "program-eligible" but did not receive a clinician response.

We propose three hypotheses related to incident-level outcome domains. By their nature, these effects will be isolated to the circumstances of the specific incident and will not pick up effects occurring because of community-level changes in the way mental health crises are handled as a whole. Rather, we anticipate that the effects we observe from this research agenda will be due to the presence of the mental health clinician on scene in collaboration with

1. One city not included in these analyses only recorded when a clinician showed up to a call for service.

co-responding police officers to provide individuals with the most effective resources possible (and avoid entering the criminal justice system or being subjected to an unnecessary involuntary psychiatric detention). We have three critical hypotheses for these analyses as follows:

Hypothesis B1
CWCRT-eligible incidents that receive a co-response will experience a lower frequency of involuntary psychiatric detentions compared to otherwise similar CWCRT-eligible incidents that do not receive a co-response due to unavailability of the requested mental health clinician (i.e., clinician is busy or off duty).

Hypothesis B2
CWCRT-eligible incidents receiving a co-response will experience a lower frequency of criminal offenses (i.e., the opening of a criminal case in response to a documented complaint, warning, citation, or arrest) registered compared to otherwise similar CWCRT-eligible incidents that do not receive a co-response due to unavailability of the requested mental health clinician (i.e., clinician is busy or off duty).

Hypothesis B3
CWCRT-eligible incidents that receive a co-response will experience a lower frequency of arrests compared to otherwise similar CWCRT-eligible incidents that do not receive a co-response due to unavailability of the requested mental health clinician (i.e., clinician is busy or off duty).

Prior to any data analysis, we preregistered both of these research agendas and their associated domains of outcomes with the Open Science Framework (https://osf.io/dbs35). Doing so ensures that our research adheres to high standards of practice in quasi-experimental research design and analysis. We share information on the preregistration in the Appendix, including information on any deviations we have made from the preregistration protocol.

Data and Methods

For both the community- and incident-level effects research agendas, we received data transfers from San Mateo County or a particular focal city by means of signed data use agreements with the relevant governing bodies. We retain emergency dispatcher call records, criminal cases, arrests, and involuntary psychiatric detention data for each focal police area in a given month from 2019 through 2023. Below, we provide details on the procedures for defining the sample and analytic strategies in each of the two research agendas.

Community-Level Research Agenda

In this research agenda, we focus on four domains of study to evaluate the effect of the CWCRT program on:

1. Frequency of program-related incidents/calls for service received by emergency dispatchers;
2. Criminal offenses related to the program's stated goals;
3. Arrests related to the program's stated goals; and
4. Involuntary psychiatric detentions.

Specifically, the data come from emergency dispatcher incidents/calls for service and criminal complaints and arrests provided by San Mateo County and/or police agencies in the cities of Belmont, Burlingame, Daly City, Foster City, Menlo Park, Pacifica, San Mateo, South San Francisco, and Redwood City.

Our focal timeframe and unit of analysis is program-related calls for service per police area in 48 police areas (26 treatment, 22 comparison) among these nine cities in San Mateo County over 54 months. This configuration resulted in 2,592 police area-month observations from January 1, 2019, through June 30, 2023—a time period representing the first year and a half of the pilot phase of CWCRT up to the last round of full data collection (December 2021–June 2023) and the three years prior to the pilot launch (January 2019–November 2021). Our DiD design strategy allows us to take advantage of our panel dataset in months surrounding implementation. In order to differentiate incidents and offenses related to CWCRT operations, we identified and coded the types of reported incidents that the CWCRT teams would most likely be sent to and offenses most likely to be reduced by the program (i.e., "CWCRT-related" incidents and offenses).

This research domain's first confirmatory outcome of interest is police area-month counts of CWCRT-related incidents. According to the CWCRT program's "Theory of Change," the program dispatches a co-response unit consisting of law enforcement officers and mental health clinicians to incidents that "involve a mental health component" (Gardner Center, 2024d). This system relies on police officer and dispatcher discretion. We code CWCRT-related calls for service based on the three most prevalent call types among CWCRT-eligible calls and CWCRT co-responses: welfare check, mental health, and community disturbance calls for service. Along with this confirmatory outcome are three exploratory outcomes, which are separated counts of each of the three most prevalent CWCRT-eligible calls: welfare check, mental health, and community disturbance call counts.

The second confirmatory outcome in this domain is program-related criminal-offense counts, which are instead differentiated by types of offenses that may be related to the program's stated goal of addressing situations involving a mental health-related crisis. We code criminal offense data as "program-related" if the offense is associated with one of the three most prevalent calls for service in CWCRT-eligible calls and responses (i.e., welfare check, mental health,

and community disturbance calls). Using these data, we first construct a simple binary indicator equal to 1 for police area-month observations from police areas that participate in CWCRT during a given month (i.e., a "static" measure of treatment).

The third confirmatory outcome is program-related arrests, which like criminal offenses are also coded as "program-related" if the arrest is associated with a welfare check, mental health, or community disturbance call for service. Using these data, we similarly construct a simple binary indicator equal to 1 for police area-month observations from police areas that participate in CWCRT during a given month (i.e., a "static" measure of treatment).

The fourth and final outcome is whether an involuntary psychiatric detention was ordered for a given incident, which by definition constitutes a program-related incident.

Descriptive statistics shown in Table 1 indicate that 19% of area-month calls in the data received the CWCRT co-responder treatment program from January 2019 through June 2023. Across the entire time period, there are an average of about 70 program-related calls for service per month. Broken down by the types of calls that comprise this confirmatory outcome, about 23 calls per month are welfare checks, 3 are mental health calls, and 44 are community disturbance calls. About 3 incidents per police area-month involved an involuntary psychiatric detention, while 3.6 calls per month in the average police area involved a CWCRT-related police case being opened, 3.3 calls resulted in a criminal offense being logged, and 1.7 program-related calls resulted in an arrest. These area-month data are also largely right-skewed and register substantial zero-count area-months to varying degrees. Specifically, 14% of area-months in the data register zero counts of CWCRT-related calls for service, 16% have zero area-month instances of welfare checks, 28% have no mental health calls, and 19% have no community disturbance calls. Greater missingness is registered for the types of actions taken as a result of program-related calls; 31% of police area-months report no instances of involuntary psychiatric detentions, about 30% record no program-related police cases or criminal offenses, and 47% record no program-related arrests.

Given the prevalence of zero count police area-months in the data, our preregistered analyses rely on recent guidance by Chen and Roth (2023), who offer suggested solutions for properly handling count data with a prevalence of zero-count cells. As a result, our preferred models for all count outcomes are conditional maximum likelihood Poisson models with robust standard errors. The specification for each outcome of interest takes the following form:

$$Y_{at} = \alpha_a + \gamma_t + \Theta S_{at} + \varepsilon_{at} \quad (1)$$

where Y_{at} is the log-linear expression for a given count outcome within police area α and time period t. The term S_{at} is a binary indicator equal to 1 only for police areas participating in the CWCRT program during time periods when the co-response program was operating. The coefficient of interest θ represents the effect of the CWCRT program conditional on fixed effects unique to each police area and to each month (i.e., α_a and γ_t, respectively). The term ε_{at} is a mean-zero error term with clustering at the police area level.

Equation 1 implies that the CWCRT program produces a single and sustained effect on each outcome of interest once the program begins in a particular police area. In order to consider whether post-treatment effects of the program vary over time, we also report the results of "semi-dynamic" DiD that allows for effects to vary uniquely and unrestrictively

in each of the first nine months of program implementation, and then the average among months thereafter. We also present the results based on versions of Equation 1 that use alternative approaches to estimation that account for high zero counts and overdispersion of the dependent variable (e.g., negative binomial conditional maximum likelihood count data specifications).

Equation 1 presents a "static" DiD specification which rests on the assumption that effects of the CWCRT program are activated once treatment begins and are constant over time. However, the nature of the program's rollout implies there may be dynamic features to the program's effects. We rely on a semi-dynamic DiD model to test for time-varying treatment effects, which unrestrictively allows for treatment effects unique to the month immediately after a precinct first participates and up to five months later:

$$Y_{am} = \alpha_a + \gamma_m + \sum_{n=0}^{10} \delta_{-n} S_{a,m-n} + \varepsilon_{am} \qquad (2)$$

In this model, the coefficients of interest are represented by δ_n, which identify the effects of CWCRT in the first month of the program (i.e., $S_{a,m-0}$) as well as the current effect of having begun one month earlier (i.e., $S_{a,m-1}$), two months earlier (i.e., $S_{a,m-2}$), and so on. We then test the equivalence of these coefficients of interest using the null hypothesis of a constant treatment effect:

$$H_{10}: \delta_0 = \delta_{-1} = \delta_{-2} \ldots = \delta_{-10}$$

We report the semi-dynamic DiD results for each outcome of interest in Table 3.

Finally, DiD specifications compare the before/after level of focal outcomes in police areas employing the CWCRT program to changes in comparison police areas where services were not yet available. For a causal interpretation of our confirmatory DiD estimates, we must rely on the crucial assumption that time-varying characteristics of comparison police areas serve as valid counterfactuals for what we would have expected to happen in treated police areas if not for the presence of the CWCRT program. Our primary approach for interrogating this key assumption of "parallel trends" in counterfactual post-treatment outcomes is by estimating event study models that unrestrictively allow for month-specific effects of CWCRT that are unique to treatment police areas. Our event study model allows us to evaluate whether there is reason to believe treatment and comparison police areas exhibit similar variation in pre-treatment trends. To the extent that this hypothesis is true, it is consistent with the parallel-trends assumption. We examine this question through event-study specifications of the following form:

$$Y_{am} = \alpha_a + \gamma_m + \sum_{\tau=1}^{12} \delta_\tau S_{a,m+\tau} + \sum_{n=0}^{10} \delta_{-n} S_{a,m-n} + \varepsilon_{am} \qquad (3)$$

This event-study specification effectively extends the semi-dynamic specification (Equation 2) to allow for fixed effects unique to each month prior to participating in CWCRT (i.e., "leads" of treatment adoption). That means the coefficients of interest are represented as δ_{-n} and δ_{τ}, which designate the "effect" for police area a in month m of participation in CWCRT n months in the future or τ months in the past. The reference category includes those never participating in CWCRT and those that are 13 or more months prior to their first participation in CWCRT. To examine the assumption of parallel trends, we test whether — *prior* to their participation in CWCRT — treatment precincts have month-to-month changes in outcomes distinct from comparison precincts:

$$H_0: \delta_{12} = \delta_{11} = \ldots \delta_1 = 0$$

We report the event-study results for each outcome of interest in Table 4.

Incident-Level Research Agenda

Our second research agenda isolates the incident-level effects of the CWCRT program on criminal offenses, number of arrests, and involuntary psychiatric detentions for all reported incidents deemed "CWCRT eligible" by a dispatcher or police officer. Eligible calls are logged regardless of whether a clinician was available to respond to the incident. Our focal timeframe and unit of analysis is CWCRT-eligible (dispatcher determined or officer) incidents/calls for service in three of the four cities participating in the CWCRT program from January 1, 2022, through December 31, 2023. These analyses exclude data from Daly City, which only reported on incidents in which CWCRT clinicians responded with officers and not on those eligible for the service when the clinician was unavailable to respond. Our fixed-effects design strategy effectively compares outcomes conditional on effects associated with police area, year, month of year, day of week, and/or hour of day.

These data come from emergency dispatcher incidents/calls for service, criminal offenses, and related arrest data provided by San Mateo County and/or police agencies in the CWCRT participating cities of San Mateo, South San Francisco, and Redwood City. These data involving adults are available from January 1, 2022, through December 31, 2023.

Descriptive statistics reported in Table 6 indicate that of the 4,117 program-eligible calls for service documented in the 24 months under study, 29% (n = 1,208) received a CWCRT co-response while 71% (n = 2,909) did not. The majority of program-eligible calls not receiving a co-response are situated in the City of San Mateo (78%), while 15% are logged in South San Francisco, and just 6% come from Redwood City. On the other hand, about a third of program-eligible calls receiving a co-response are in Redwood City, 45% are in the City of San Mateo, and 22% are in South San Francisco. Program-eligible calls not receiving a response are fairly evenly dispersed throughout the week, with 16% of all eligible calls not receiving a response recorded on Sundays, between 12% and 13% on each day from Monday through Thursday, and 17% on both Friday and Saturday. On the other hand, 18% of co-responses occurred on a Monday, while 24% were logged on Tuesdays, 23% on Wednesdays, 25% on Thursdays, and 9% on Fridays. No responses were logged on Saturdays and Sundays, when clinicians were not on duty in any of the three cities.

Similar descriptive patterns emerge in Table 6 among hours of the day. We found that 11% of eligible calls not receiving a response occur between midnight and 4 a.m., 8% occur between 4 a.m. and 8 a.m., 18% occur between 8 a.m. and noon, 17 percent are reported between noon and 4 p.m., 25% between 4 p.m. and 8 p.m., and 21% between 8 p.m. and midnight. Actual program co-responses occurred mostly between 8 a.m. and 4 p.m. (84%) when clinicians were typically on duty. No responses occurred between 8 p.m. and 4 a.m. By month and year, roughly 31% of CWCRT-eligible calls came in from December 2021 through May 2022, 29% from June 2022 through November 2022, 20% from December 2022 through May 2023, and 20% from June 2023 through December 2023, regardless of whether a co-response occurred or not. Finally, throughout this time period about 17% of CWCRT-eligible calls are logged as implementing involuntary psychiatric detentions, while just 5% recorded a police case being opened, 4% documented a criminal offense, and 3% resulted in an arrest, regardless of whether a clinician responded or not.

Our main confirmatory analysis is a fixed-effects analysis that effectively compares preregistered outcomes across CWCRT-eligible incidents that did and did not have a co-responding clinician (i.e., due to a clinician being unavailable or off duty). Our baseline specification expresses the outcome for incident i, Y_{ichdt}, as a function of unrestrictive fixed effects for each city, c; each hour of the day, h; each day of the week, d; and each month-year combination, t:

$$Y_{ichdt} = \alpha_c + \gamma_h + \delta_d + \eta_t + \beta C_{ichdt} + \varepsilon_{ichdt} \qquad (4)$$

Our coefficient of interest, β, identifies the effect of having a co-responder dispatched to the incident (i.e., C_{ichdt}) conditional on these fixed effects. We explore the robustness of results based on this design in several ways. First, we evaluate estimates of β in specifications saturated with increasingly unrestrictive sets of fixed effects (e.g., for each unique city-month-year combination and each interaction of the day of the week and the hour of the day). Second, we consider the reweighting approach introduced by Gibbons et al. (2019) to assess the potential bias that can occur when there is variation in the number of observations and the conditional variance of treatment within cells defined by fixed effects. Third, though our main results rely on robust standard errors, we also consider inference under different forms of clustering as well as randomization inference.

Results

Community-Level Analyses

We begin with considering the total effects of the CWCRT pilot program by analyzing counts of program-related calls for service, involuntary psychiatric detentions, police cases, criminal offenses, and arrests in each of 48 focal police areas in months from January 1, 2019, through June 30, 2023. All main total-effects analyses rely on conditional maximum likelihood Poisson models with bootstrapped standard errors blocked at the police area level. These results are presented as the specific average area-month percent change in the prevalence of each outcome due to the program's implementation.

Table 2 displays the estimated treatment effects of the CWCRT program on the one-time change in each of the outcomes of interest. The first column is the program's effect on the prevalence of program-related calls for service (i.e., a combined count of welfare check, mental health, and community disturbance calls in each police area per month), suggesting null treatment effects (b = -0.02, SE = 0.04, p = 0.64). Static DiD models also suggest null effects on welfare check calls (b = 0.07, SE = 0.06, p = 0.25) and community disturbances (b = -0.03, SE = 0.05, p = 0.52). However, once the CWCRT program became available in participating police areas, there is a statistically significant effect of the program on mental health calls registered in those treated police areas (b = -0.19, SE = 0.09, p = 0.03), amounting to a (1 - exp[-0.19] =) 17% reduction in these calls. Likewise, treated police area-months also experienced a statistically significant decrease in the number of involuntary psychiatric detentions implemented due to program implementation (b = -0.18, SE = 0.07, p = 0.02), amounting to a (1 - exp[-0.19] =) 16% reduction in involuntary psychiatric detentions. Finally, we observe negatively trending estimates for program-related counts of police cases (b = -0.35, SE = 0.21, p = 0.10), criminal offenses (b = -0.20, SE = 0.20, p = 0.32), and arrests (b = -0.26, SE = 0.20, p = 0.19) in treated area-months, but these outcomes appear to be estimated imprecisely and are thus not statistically significant at the p < 0.05 level.

Semi-dynamic DiD results reported in Table 3 suggest post-treatment trends in the prevalence of combined CWCRT-related calls for service (p = 0.04), welfare check calls (p = 0.04), and community disturbance calls (p < 0.01). However, Table 4 results suggest likely violations of the crucial "parallel trends" assumption required to credibly interpret causal effects for combined CWCRT-related calls (p = 0.01), welfare check calls (p = 0.06), community disturbance calls (p = 0.04), and program-related police cases (p < 0.01), criminal offenses (p < 0.01), and arrests (p = 0.02). On the other hand, we cannot reject the null hypotheses that these results do not violate the parallel trends assumption for mental health calls (p = 0.44) and involuntary psychiatric detentions (p = 0.57), allowing for credible causal interpretations of these two outcomes. Thus, from Tables 2 and 3 we can infer that the prevalence of both mental health calls and involuntary psychiatric detentions dropped most drastically in the first five months of CWCRT program implementation (i.e., drops as large as 53% in these months), followed by smaller negative effects in the subsequent months of program implementation (i.e., between 10% and 30% decreases in months thereafter).

Finally, the many robustness checks reported in Table 5 suggest remarkably consistent negative total effects of the CWCRT program on involuntary psychiatric detentions and mental health calls across many different model specifications, including our preferred conditional maximum likelihood Poisson model with robust standard errors; a conditional maximum likelihood negative binomial model with robust standard errors; conditional maximum likelihood Poisson and negative binomial models with bootstrapped standard errors blocked within police areas; maximum likelihood Poisson and negative binomial models with robust standard errors; and maximum likelihood Poisson and negative binomial models with errors clustered within police areas.

Incident-Level Analyses

Our incident-level analyses consider fixed effects differentiating program-eligible incidents receiving a CWCRT response versus those eligible but not receiving a response (i.e., due to a clinician being unavailable or off duty) on four dimensions of interest: city, year-month sequence, day of the week, and hour of day (Gibbons et al., 2019). Our main confirmatory analysis considers one-way interactions of these four dimensions; however, we test the robustness of these results in several ways. We first introduce two-way interactions among these dimensions (e.g., the number of calls in Redwood City on Fridays, or the number of calls occurring at 3 p.m. in October 2022). Next, we introduce three-way interactions among the four dimensions — for example, considering the number of calls occurring in Redwood City at 3 p.m. on Fridays.

In introducing these increasingly restrictive parameters, we lose cells of data that do not have valid comparison cases (i.e., we lose "common support" in our analysis). Table 7 displays the number of observations we are able to observe in each of these model specifications. The first row includes all information in the sample; among the 4,117 calls observed, 681 resulted in an involuntary psychiatric detention, 209 had a police case opened, 183 included a criminal offense being logged, and 130 ended in arrest. When considering all one-way interactions among the four dimensions of interest, the common support of the sample drops to 3,221 calls, 496 of which result in an involuntary psychiatric detention, 157 in a police case opened, 139 in a criminal offense logged, and 97 in an arrest. Including two-way interactions reduces the sample of common support to 2,011 calls (i.e., about half of the full sample of calls). Finally, including the three-way interactions reduces the sample of common support to just 621 calls (i.e., just 15% of all calls observed), only 69 of which result in an involuntary psychiatric detention, 35 in a police case opened, 29 in a criminal offense logged, and 17 in arrest. These results suggest modest common support for the two-way fixed-effects interaction model and poor common support for the three-way fixed-effects interactions model. Results from these increasingly saturated fixed-effects model specifications that build from fixed effects identified in Equation 4 are reported in Tables 8 through 11.

Table 8 reports the incident-level effects of a CWCRT co-response on the probability the call will result in an involuntary psychiatric detention. The first column represents the model identified in Equation 4, with fixed-effects interactions among police agencies, month-year sequence, day of week, and hour of day. This model suggests that a CWCRT co-response is associated with a 12-percentage point reduction in the probability that the incident will result in an involuntary psychiatric detention ($p < 0.01$). This model explains about 33% variance in the dependent variable. In column two we add to the base model all two-way interactions among a police agency, month-year sequence, day of week, and hour of day fixed effects. This model suggests a co-response is associated with an 8-percentage point reduction in the probability of an involuntary emergency psychiatric hold ($p < 0.01$), explaining a third of the association from model 1 and 52% of the variance in the dependent variable.

Finally, the third column in Table 8 includes three-way interactions among a city, day of week, hour of day, and year-month sequence. This more saturated model (which drops all but 621 of the 4,117 calls and all but 69 of the nearly 500 involuntary psychiatric detentions in the data) explains 86% of the variance in the dependent variable but suggests a CWCRT co-response has no association with the probability a call will result in an involuntary psychiatric detention (b = 0.012, SE = 0.029, $p = 0.66$). However, as Table 7 suggests, this more saturated model lacks substantial common support for all the variables of interest. All other outcomes register near-zero results across different model specifications (Tables 8-11).

Finally, applying guidance from Gibbons et al. (2019), we perform a crucial robustness check to evaluate the degree to which the fixed effects models we employ apply tacit weights to cells with greater instances of CWCRT-eligible calls for service. Specifically, we compare our confirmatory and exploratory OLS fixed-effects estimators with regression-weighted estimators (RWE) that correct for the potential implicit weighting of the OLS models. Table 12 provides chi-squared statistics and p-values testing the null hypothesis that the OLS and RWE estimates of a given model specification are equal. The first column presents tests of our confirmatory analyses. At the p <.05 level, chi-squared tests fail to reject the null hypothesis that the OLS and RWE results are similar in size for police cases (p = 0.28), criminal offenses (p = 0.27), and arrests (p = 0.62). In contrast, although the test does not reject the null hypothesis for involuntary psychiatric detentions, the evidence in favor of the null hypothesis in this confirmatory model is weak (p = 0.07). The models including two-way FE interactions provide stronger evidence of failing to reject the null hypothesis, particularly for involuntary psychiatric detentions (p = 0.96), and the three-way FE interaction models all fail to reject the null hypothesis of equivalent OLS and RWE estimators.

Conclusion

Confirmed across two independent, preregistered, quasi-experimental design agendas, our results suggest that the CWCRT program led to reductions in police-initiated involuntary psychiatric detentions among participating communities. To put these results into perspective, in the year prior to program adoption, the four cities participating in CWCRT together documented 1,011 instances of involuntary psychiatric detentions initiated by police. Had the CWCRT program been available in that year, our estimates of a (1 - exp[-0.18] =) 16% reduction would anticipate [1,011 * 0.165 * 2 =] 333 fewer involuntary psychiatric detentions over the two pilot years than those reported by the four cities (or [333 / 4 =] 83 fewer detentions per agency over the two pilot years).

At the incident level, the program's effects amount to an estimated 7.8 to 11.5 percentage point reduction in the probability of implementing an involuntary psychiatric detention during a program-eligible call for service. Given the 1,208 calls among the three cities in the analysis involving a co-response during the two-year CWCRT pilot period, we estimate between (1,208 * 0.078 =) 94 and (1,208 * 0.115 =) 139 fewer involuntary psychiatric detentions occurred in these three cities over the first two years of the pilot (or between [94 / 3 =] 31 and [139 / 3 =] 46 fewer detentions per agency over the two pilot years).

We note that an incident-level analysis (i.e., an assessment conditional on a CWCRT-eligible incident occurring) results in a smaller estimated impact on involuntary psychiatric detentions than our community-wide analysis. These comparative results provide indirect evidence consistent with the hypothesis that the CWCRT program both reduced the probability of a given incident resulting in an involuntary psychiatric detention and reduced the likelihood of such incidents happening again during the CWCRT pilot period. In other words, these comparative impact estimates suggest that the presence of co-responders improved the continuum of care in a manner that likely reduced repeated need for police intervention and emergency psychiatric services.

References

Butts, K. (2023). Geographic difference-in-discontinuities. *Applied Economics Letters*, 30(5), 615-619.

Chen, J., & Roth, J. (2023). Logs with zeros? Some problems and solutions. *The Quarterly Journal of Economics*, 139(2), 891–936.

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

Gardner Center. (2024d). *Research brief: Pilot program theory of change, Community Wellness and Crisis Response Team of San Mateo County, California*. Stanford, CA: The John W. Gardner Center for Youth and Their Communities at Stanford University.

Gibbons, C. E., Suárez Serrato, J. C., & **Urbancic, M. B.** (2019). Broken or fixed effects? *Journal of Econometric Methods*, 8(1), 1–12.

Keele, L. J., & Titiunik, R. (2015). Geographic boundaries as regression discontinuities. *Political Analysis*, 23(1), 127-155.

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

Table 1. Descriptive Statistics, Area-Month Panel Data

Variable	Mean	SD	Min	Max	% with Zero Counts
Co-Responder Treatment	0.19	0.39	0	1	—
CWCRT-Related Calls for Service	69.56	55.29	0	271	0.14
Welfare Check Calls	22.57	22.38	0	175	0.16
Mental Health Calls	3.30	3.43	0	19	0.28
Community Disturbance Calls	43.69	38.56	0	185	0.19
Involuntary Psychiatric Detentions	2.84	3.09	0	17	0.31
CWCRT-Related Police Cases	3.61	4.46	0	34	0.29
CWCRT-Related Criminal Offenses	3.33	4.09	0	26	0.30
CWCRT-Related Arrests	1.71	2.44	0	17	0.47

Note: Sample consists of a balanced panel of 48 police areas over 54 months (N = 2,592). “CWCRT-Related” is defined as all calls coded by dispatchers as welfare checks, mental health calls, or disturbances.

Table 2. Static Difference-in-Differences Analyses, Area-Month Panel Data

Independent Variable	CWCRT-Related Calls	Welfare Check Calls	Mental Health Calls	Commun. Disturbance Calls	Involuntary Psychiatric Detentions	CWCRT-Related Police Cases Opened	CWCRT-Related Criminal Offenses Recorded	CWCRT-Related Arrests
Treatment	-0.02	0.07	-0.19**	-0.03	-0.18**	-0.35	-0.20	-0.26
	(0.04)	(0.06)	(0.09)	(0.05)	(0.07)	(0.21)	(0.20)	(0.20)

Note: The analytic sample is a balanced panel of 48 police areas observed in each of 54 months (N = 2,592). The reported estimates are based on conditional maximum likelihood (CML) Poisson specifications that include area fixed effects and month-year fixed effects. The effective sample sizes vary because the conditional maximum likelihood function excludes police areas without within-area variation in the given outcome. Bootstrapped standard errors in parentheses are blocked at the police area level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3. Semi-Dynamic Difference-in-Differences Analyses, Area-Month Panel Data

Independent Variable	CWCRT-Related Calls	Welfare Check Calls	Mental Health Calls	Community Disturbance Calls	Involuntary Psychiatric Detentions	CWCRT-Related Police Cases Opened	CWCRT-Related Criminal Offenses Recorded	CWCRT-Related Arrests
Initial treatment	-0.06	-0.05	0.01	-0.03	0.08	0.00	-0.03	-0.05
	(0.08)	(0.08)	(0.25)	(0.12)	(0.26)	(0.16)	(0.18)	(0.29)
1-month lead	-0.10*	0.11	-0.30*	-0.19***	-0.29	-0.14	-0.10	-0.06
	(0.06)	(0.10)	(0.18)	(0.07)	(0.20)	(0.19)	(0.19)	(2.02)
2-month lead	-0.09**	-0.06	-0.04	-0.11**	-0.09	-0.49	-0.33	-0.38
	(0.04)	(0.07)	(0.25)	(0.05)	(0.24)	(0.38)	(0.33)	(0.32)
3-month lead	-0.07	-0.09	-0.49***	0.02	-0.53***	-0.51	-0.26	-0.06
	(0.05)	(0.07)	(0.18)	(0.07)	(0.19)	(0.32)	(0.32)	(0.35)
4-month lead	-0.13*	-0.14	-0.38**	-0.04	-0.38**	-0.32	-0.15	0.00
	(0.07)	(0.10)	(0.17)	(0.08)	(0.16)	(0.39)	(0.38)	(2.13)
5-month lead	-0.05	0.01	-0.24	-0.06	-0.17	-0.40**	-0.30*	-0.66**
	(0.05)	(0.10)	(0.18)	(0.07)	(0.20)	(0.19)	(0.17)	(0.32)
6-month lead	-0.01	0.05	-0.16	-0.02	-0.19	-0.22	-0.04	-0.10
	(0.06)	(0.10)	(0.20)	(0.08)	(0.18)	(0.29)	(0.29)	(0.37)
7-month lead	0.12	0.23*	-0.20	0.07	-0.19	-0.21	-0.03	-0.31
	(0.08)	(0.13)	(0.20)	(0.07)	(0.17)	(0.31)	(0.30)	(0.40)
8-month lead	0.05	0.18	-0.15	-0.01	-0.19	-0.13	-0.05	-0.34
	(0.06)	(0.11)	(0.27)	(0.07)	(0.32)	(0.28)	(0.25)	(0.32)
9-month lead	-0.02	0.09	-0.29	-0.06	-0.30	-0.06	0.06	0.14
	(0.06)	(0.10)	(0.22)	(0.07)	(0.20)	(0.32)	(0.30)	(0.41)
10-month +	-0.02	0.07	-0.14	-0.03	-0.10	-0.38*	-0.19	-0.27
	(0.05)	(0.07)	(0.11)	(0.06)	(0.10)	(0.21)	(0.21)	(0.25)
p-value (lags)	**0.04**	**0.04**	**0.12**	**< 0.01**	**0.13**	**0.30**	**0.43**	**0.67**

Note: The analytic sample is a balanced panel of 48 police areas observed in each of 54 months (N = 2,592). The reported estimates are based on conditional maximum likelihood (CML) Poisson specifications that include area fixed effects and month-year fixed effects. The effective sample sizes vary because the conditional maximum likelihood function excludes police areas without within-area variation in the given outcome. Bootstrapped standard errors in parentheses are blocked at the police area level. *** p < 0.01, ** p < 0.05, * p < 0.10.

Table 4. Event Studies, Area-Month Panel Data

Independent Variable	CWCRT-Related Calls	Welfare Check Calls	Mental Health Calls	Community Disturbance Calls	Involuntary Psychiatric Detentions	Related Police Cases Opened	Related Criminal Offenses Recorded	Related Arrests
12-month lag	0.05	0.14	-0.01	0.00	0.05	-0.10	-0.03	0.30
	(0.07)	(0.11)	(0.22)	(0.08)	(0.21)	(0.29)	(0.35)	(2.86)
11-month lag	0.01	0.01	-0.04	0.04	0.03	-0.25	-0.26	-1.04**
	(0.06)	(0.08)	(0.23)	(0.10)	(0.19)	(0.44)	(0.44)	(0.42)
10-month lag	0.11	0.12	0.13	0.15	0.20	-0.60***	-0.59**	-1.19***
	(0.08)	(0.10)	(0.18)	(0.09)	(0.21)	(0.21)	(0.23)	(0.37)
9-month lag	0.02	0.11	0.26	-0.05	0.25	0.17	0.18	-0.32
	(0.07)	(0.07)	(0.18)	(0.08)	(0.20)	(0.24)	(0.25)	(2.20)
8-month lag	0.03	0.01	-0.02	0.07	0.06	0.20	0.21	-0.11
	(0.06)	(0.07)	(0.23)	(0.07)	(0.18)	(0.26)	(0.25)	(0.54)
7-month lag	0.00	-0.05	-0.05	0.10	-0.05	0.57**	0.58**	-0.15
	(0.06)	(0.08)	(0.28)	(0.07)	(0.27)	(0.28)	(0.29)	(1.64)
6-month lag	0.02	0.10	-0.13	-0.01	-0.05	0.06	0.04	-0.29
	(0.06)	(0.09)	(0.23)	(0.07)	(0.27)	(0.18)	(0.20)	(0.30)
5-month lag	0.01	-0.05	0.16	0.05	0.23	0.27	0.25	0.15
	(0.08)	(0.09)	(0.28)	(0.08)	(0.31)	(0.24)	(0.25)	(0.29)
4-month lag	-0.07	-0.03	0.01	-0.10	0.08	-0.38	-0.34	0.22
	(0.05)	(0.08)	(0.21)	(0.07)	(0.19)	(0.33)	(0.36)	(3.12)
3-month lag	0.00	-0.03	0.32	0.03	0.35*	0.18	0.11	0.39
	(0.05)	(0.08)	(0.25)	(0.06)	(0.21)	(0.23)	(0.24)	(2.86)
2-month lag	-0.13**	-0.12	-0.14	-0.09	-0.10	0.23	0.19	-0.44
	(0.06)	(0.08)	(0.21)	(0.10)	(0.21)	(0.19)	(0.19)	(0.27)
1-month lag	-0.12	-0.15	-0.18	-0.02	-0.10	-0.13	-0.16	-0.30
	(0.09)	(0.11)	(0.24)	(0.09)	(0.24)	(0.26)	(0.27)	(0.47)
Initial treatment	-0.06	-0.05	0.01	-0.02	0.10	0.01	-0.02	-0.13
	(0.09)	(0.09)	(0.26)	(0.12)	(0.26)	(0.17)	(0.18)	(0.29)
1-month lead	-0.11	0.11	-0.30	-0.19**	-0.27	-0.14	-0.10	-0.14
	(0.07)	(0.10)	(0.19)	(0.08)	(0.21)	(0.21)	(0.21)	(2.09)
2-month lead	-0.09*	-0.06	-0.03	-0.11*	-0.07	-0.48	-0.33	-0.46
	(0.05)	(0.08)	(0.26)	(0.06)	(0.24)	(0.38)	(0.33)	(0.33)
3-month lead	-0.07	-0.09	-0.49***	0.02	-0.50***	-0.51	-0.25	-0.13
	(0.05)	(0.07)	(0.18)	(0.08)	(0.19)	(0.32)	(0.32)	(0.36)
4-month lead	-0.13*	-0.14	-0.37**	-0.04	-0.36**	-0.31	-0.15	-0.08
	(0.08)	(0.11)	(0.18)	(0.09)	(0.17)	(0.38)	(0.38)	(2.17)
5-month lead	-0.05	0.01	-0.24	-0.06	-0.14	-0.40**	-0.30*	-0.74**
	(0.05)	(0.10)	(0.20)	(0.07)	(0.21)	(0.19)	(0.17)	(0.33)
6-month lead	-0.02	0.05	-0.15	-0.02	-0.16	-0.21	-0.04	-0.18
	(0.06)	(0.10)	(0.22)	(0.09)	(0.19)	(0.28)	(0.28)	(0.37)
7-month lead	0.11	0.23*	-0.19	0.07	-0.17	-0.20	-0.03	-0.39
	(0.08)	(0.13)	(0.21)	(0.08)	(0.18)	(0.31)	(0.29)	(0.41)
8-month lead	0.05	0.18	-0.14	0.00	-0.16	-0.12	-0.05	-0.42
	(0.06)	(0.11)	(0.27)	(0.07)	(0.32)	(0.28)	(0.25)	(0.34)
9-month lead	-0.02	0.09	-0.28	-0.05	-0.27	-0.06	0.06	0.06
	(0.07)	(0.09)	(0.22)	(0.08)	(0.20)	(0.32)	(0.30)	(0.41)
10-month +	-0.02	0.07	-0.13	-0.03	-0.08	-0.37*	-0.18	-0.35
	(0.05)	(0.07)	(0.13)	(0.06)	(0.10)	(0.21)	(0.20)	(0.25)
p-value (leads)	0.01	0.06	0.44	0.04	0.57	< 0.01	< 0.01	0.02
p-value (lags)	0.04	0.04	0.12	< 0.01	0.13	0.30	0.43	0.67

Note: The analytic sample is a balanced panel of 48 police areas observed in each of 54 months (N = 2,592). The reported estimates are based on conditional maximum likelihood (CML) Poisson specifications that include area fixed effects and month-year fixed effects. The effective sample sizes vary because the conditional maximum likelihood function excludes police areas without within-area variation in the given outcome. Bootstrapped standard errors in parentheses are blocked at the police area level. *** p < 0.01, ** p < 0.05, * p < 0.10.

Table 5. Robustness Checks, Area-Month Panel Data

	Dependent Variable					
	Involuntary Psychiatric Detentions			Mental Health Calls		
Model Specification	Est.	SE	p	Est.	SE	p
Bootstrapped CML Poisson	-0.18	0.07	0.02	-0.19	0.09	0.03
Bootstrapped CML NB	-0.17	0.07	0.02	-0.19	0.09	0.04
CML Poisson	-0.18	0.07	0.02	-0.19	0.08	0.02
CML NB	-0.17	0.06	< 0.01	-0.19	0.06	< 0.01
ML Poisson Robust	-0.18	0.06	< 0.01	-0.19	0.06	< 0.01
ML NB Robust	-0.18	0.06	< 0.01	-0.19	0.06	< 0.01
ML Poisson Clustered SEs	-0.18	0.08	0.02	-0.19	0.08	0.02
ML NB Clustered SEs	-0.18	0.08	0.02	-0.19	0.08	0.02

Note: Analytic sample consists of a balanced panel of 48 police areas over 54 months (N = 2,592). NB = Negative Binomial; CML = Conditional Maximum Likelihood; ML = Maximum Likelihood. “Clustered SEs” refers to standard errors clustering at the police area level.

Table 6. Descriptive Statistics, Incident-Level Data

Variable	Calls Without Co-Response (N = 2,909)	Calls with Co-Response N = 1,208)	Total Calls (N = 4,117)
Redwood City	0.06	0.34	0.14
San Mateo	0.78	0.45	0.68
South San Francisco	0.15	0.22	0.17
Sunday	0.16	0.00	0.12
Monday	0.13	0.18	0.14
Tuesday	0.13	0.24	0.16
Wednesday	0.12	0.23	0.15
Thursday	0.13	0.25	0.17
Friday	0.17	0.09	0.15
Saturday	0.17	0.00	0.12
12:00am - 3:59am	0.11	0.00	0.08
4:00am – 7:59am	0.08	0.02	0.06
8:00am – 11:59am	0.18	0.35	0.23
12:00pm – 3:59pm	0.17	0.49	0.27
4:00pm – 7:59pm	0.25	0.14	0.21
8:00pm – 11:59pm	0.21	0.00	0.15
Dec 2021 – May 2022	0.30	0.33	0.31
June 2022 – Nov 2022	0.31	0.25	0.29
Dec 2022 – May 2023	0.20	0.21	0.20
June 2023 – Dec 2023	0.19	0.22	0.20
Involuntary Psychiatric Detentions	0.17	0.16	0.17
Police Case Opened	0.05	0.05	0.05
Criminal Offenses Logged	0.04	0.04	0.04
Arrests	0.03	0.03	0.03

Note: Analytic sample includes 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. Daly City did not collect data on calls not receiving a response and is thus not included in these analyses.

Table 7. Counts of All Calls and Outcomes across Three Incident-Level Models

Independent Variable	All Calls	Involuntary Psychiatric Detention	Cases	Criminal Offenses	Arrests
One-Way FE	3,221	496	157	139	97
Two-FE Interactions	2,011	283	101	87	65
Three-Way FE Interactions	621	69	35	29	17

Note: Common support is based on a sample of 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. FE = Fixed effects. Fixed effects include binary variables for each city, day of the week, hour of the day, and year-month in the data. "Two-way FE Interactions" includes all two-way interactions among the four fixed-effects categories and "Three-Way FE Interactions" includes all three-way interactions among the four fixed-effects categories. Daly City did not collect data on calls not receiving a response and is thus not included in these analyses.

Table 8. Estimated Effects of Co-Response on Probability of Involuntary Psychiatric Detention, Incident-Level Data

Independent Variable	(1)	(2)	(3)
Co-Response Initiated	-0.115	-0.078	0.012
	(0.021)	(0.019)	(0.029)
Agency FE	Yes	Yes	Yes
Month-Year Sequence FE	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Hour of Day FE	Yes	Yes	Yes
All Two-Way Interactions	No	Yes	Yes
All Three-Way Interactions	No	No	Yes
R^2	0.327	0.522	0.880

Note: Analytic sample includes 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. The dependent variable is a binary indicator of whether a call for service resulted in an involuntary psychiatric detention. The common support for model 1 is 3,221 calls (496 IPDs); for model 2 is 2,011 calls (283 IPDs); and for model 3 is 621 calls (69 IPDs). Daly City did not collect data on calls not receiving a response and is thus not included in these analyses. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 9. Estimated Effects of Co-Response on Probability of a Police Case being Opened, Incident-Level Data

Independent Variable	(1)	(2)	(3)
Co-Response Initiated	0.006	-0.002	0.012
	(0.011)	(0.013)	(0.026)
Agency FE	Yes	Yes	Yes
Month-Year Sequence FE	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Hour of Day FE	Yes	Yes	Yes
All Two-Way Interactions	No	Yes	Yes
All Three-Way Interactions	No	No	Yes
R^2	0.017	0.249	0.741

Note: Analytic sample includes 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. The dependent variable is a binary indicator of whether a call for service escalated to a police case being opened as a result of the call for service. The common support for model 1 is 3,221 calls (157 cases); for model 2 is 2,011 calls (101 cases); and for model 3 is 621 calls (35 cases). Daly City did not collect data on calls not receiving a response and is thus not included in these analyses. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 10. Estimated Effects of Co-Response on Probability of a Criminal Offense being Logged, Incident-Level Data

Independent Variable	(1)	(2)	(3)
Co-Response Initiated	0.008	-0.001	0.012
	(0.01)	(0.01)	(0.02)
Agency FE	Yes	Yes	Yes
Month-Year Sequence FE	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Hour of Day FE	Yes	Yes	Yes
All Two-Way Interactions	No	Yes	Yes
All Three-Way Interactions	No	No	Yes
R^2	0.017	0.244	0.746

Note: Analytic sample includes 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. The dependent variable is a binary indicator of whether a call for service escalated to a criminal offense being documented. The common support for model 1 is 3,221 calls (139 offenses); for model 2 is 2,011 calls (187 offenses); and for model 3 is 621 calls (29 offenses). Daly City did not collect data on calls not receiving a response and is thus not included in these analyses. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 11. Estimated Effects of Co-Response on Probability of an Arrest being Made, Incident-Level Data

Independent Variable	(1)	(2)	(3)
Co-Response Initiated	0.001	-0.002	0.000
	(0.01)	(0.01)	(0.02)
Agency FE	Yes	Yes	Yes
Month-Year Sequence FE	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Hour of Day FE	Yes	Yes	Yes
All Two-Way Interactions	No	Yes	Yes
All Three-Way Interactions	No	No	Yes
R^2	0.016	0.246	0.768

Note: Analytic sample includes 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. The dependent variable is a binary indicator of whether a call for service escalated to an arrest being made. The common support for model 1 is 3,221 calls (97 arrests); for model 2 is 2,011 calls (65 arrests); and for model 3 is 621 calls (17 arrests). Daly City did not collect data on calls not receiving a response and is thus not included in these analyses. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 12. Comparisons of Incident-Level OLS Fixed Effects and Regression-Weighted Estimators Among CWCRT-Eligible Calls for Service

	One-Way FE Interactions		Two-Way FE Interactions		Three-Way FE Interactions	
Independent Variable	χ^2	*p*	χ^2	*p*	χ^2	*p*
Involuntary Psychiatric Detentions	3.41	0.07	< 0.01	0.96	< 0.01	0.99
Police Cases	1.15	0.28	0.37	0.54	0.16	0.69
Criminal Offenses	1.22	0.27	0.20	0.65	0.21	0.65
Arrests	0.24	0.62	0.17	0.68	<0.01	0.99

Note: Sample includes 4,117 CWCRT-eligible calls for service in three participating cities over 24 months. The dependent variable is a binary indicator of whether a call for service escalated to an involuntary psychiatric detention, a police case being opened, a criminal offense logged, or an arrest. Models control for agency; month-year sequence; day of week; and hour of day fixed effects; all two-way interactions of those fixed effects; and all agency-month-day three-way interactions. Daly City did not collect data on calls not receiving a response and is thus not included in these analyses. Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Appendix

Preregistration Protocol

Evaluation of the Programmatic Effects of San Mateo County's Community Wellness and Crisis Response Team Program on Related Calls for Service, Arrests, and Emergency Psychiatric Holds

Repository: Open Science Framework (https://osf.io/dbs35)

Description

The Community Wellness and Crisis Response Team (CWCRT) program is a "co-responder" initiative in which trained mental-health specialists accompany police officers on emergency calls related to nonviolent behavioral-health issues (e.g., mental-health crises, welfare checks, and substance abuse). This emerging programmatic design is a distinct alternative to the standard approach in which police officers respond to such calls for service alone. The CWCRT initiative is situated within several socioeconomically and demographically diverse communities in San Mateo County, California. Operations began at the end of December 2021 as a partnership between the San Mateo County Behavioral Health and Recovery Services (BHRS), StarVista (a community non-profit organization focused on counseling and crisis-prevention programs), and the police departments and 911 dispatchers in four pilot municipalities (i.e., Daly City, Redwood City, San Mateo, and South San Francisco). We also have contemporaneous data from six neighboring comparison cities not involved in the CWCRT pilot program.

We rely on two overarching research agendas to assess two different types of effects of the CWCRT program. The first research agenda seeks to recover the CWCRT program's total-effects, which include combined programmatic effects both at the moment of emergency response (i.e., the incident-level effects) and due to broad changes in the communities served over time (i.e., community level effects related to lowering the incidence of untreated mental-health in the community). The total-effects research agenda relies on police-area-by-month panel data on outcomes (e.g., program-related calls for service, criminal offenses, and psychiatric "5150" holds) to estimate the impact of the pilot program. This "difference-in-differences" (DD) approach effectively compares the change in outcomes among treated police areas to the contemporaneous change among untreated police areas. A key assumption of DD models is that untreated potential outcomes between the two groups proceed in parallel before exposure to the policy or program shock.

The second research agenda focuses only on data from treatment cities during the months when the CWCRT program is active. This agenda seeks to identify the immediate effect of a co-responder on proximate, incident-level outcomes (e.g., criminal offenses, arrests, 5150 holds). This approach relies on a fixed-effects design applied to an analytical sample of incidents in which a dispatcher or police officer indicated a co-responder was relevant. Conditioning on fixed effects unique to location, day of the week, hour of the day, and year-month, this approach compares outcomes among eligible incidents where a co-responder responded to eligible incidents where a co-responder did not (e.g., due to capacity constraints).

The following is our detailed pre-registration plan, filed on March 10, 2024 prior to any data analysis related to the study.

A. Study Information

1. Hypotheses

Regarding our total-effects research agenda, we consider study hypotheses differentiated by four separate outcome domains:

Domain A1: Program-related calls-for-service. Police areas participating in the CWCRT program will experience reduced frequency of program-related calls-for-service.

Domain A2: Program-related criminal offenses. Police areas participating in the CWCRT program will experience reduced frequency of criminal offenses (i.e., the opening of a criminal case in response to a documented complaint, warning, citation, or arrest) most related to CWCRT's programmatic focus (i.e., individuals in mental health distress).

Domain A3: Program-related arrests. Police areas participating in the CWCRT program will experience reduced frequency of arrests most related to CWCRT's programmatic focus (i.e., individuals in mental health distress).

Domain A4: 5150 holds. Police areas participating in the CWCRT program will experience reduced frequency of undergoing a 72-hour ("5150") emergency psychiatric holds.

For our isolated incident-level effects research agenda, we consider study hypotheses differentiated by three separate outcome domains:

Domain B1: Program-related criminal offenses. CWCRT-eligible incidents that receive a co-response will experience a lower frequency of criminal offenses (i.e., the opening of a criminal case in response to a documented complaint, warning, citation, or arrest) registered compared to otherwise similar CWCRT-eligible incidents that do not receive a co-response due to unavailability of the requested mental health clinician (i.e., clinician is busy or off duty).

Domain B2: Program-related arrests. CWCRT-eligible incidents that receive a co-response will experience a lower frequency of arrests compared to otherwise similar CWCRT-eligible incidents that do not receive a co-response due to unavailability of the requested mental health clinician (i.e., clinician is busy or off duty).

Domain B3: 5150 emergency psychiatric holds. CWCRT-eligible incidents that receive a co-response will experience a lower frequency of arrests compared to otherwise similar CWCRT-eligible incidents that do not receive a co-response due to unavailability of the requested mental health clinician (i.e., clinician is busy or off duty).

B. Design Plan

1. Study type

Observational Study - Data are collected from study subjects that are not randomly assigned to a treatment. This includes surveys, "natural experiments," and regression discontinuity designs.

2. Blinding

   No blinding is involved in this study.

3. Study design

   Regarding our total-effects research agenda, we focus on four domains of study to evaluate the effect of the CWCRT program on (1) frequency of program-related incidents/calls-for-service received by emergency dispatchers, and (2) criminal offenses related to the program's stated goals, (3) arrests related to the program's stated goals, and (4) 5150 emergency holds. Our focal timeframe and unit of analysis is program-related calls for service per police area in 48 police areas (26 treatment, 22 comparison) among nine cities in San Mateo County, CA, in a given month from January 1, 2019 through June 30, 2023. This time period represents the first year and a half of the pilot phase of CWCRT up to the last round of data collection among comparison cities (December 2021-June 2023) and the three years prior to the pilot beginning (January 2019-November 2021). Our DD design strategy allows us to take advantage of our panel dataset in months surrounding implementation.

   With respect to our isolated incident-level effects research agenda, we focus on incident-level data for the number of criminal offenses, number of arrests and 5150 holds for all reported incidents deemed "CWCRT eligible" by a dispatcher or police officer, regardless of whether a clinician was available to respond to the incident. Our focal timeframe and unit of analysis is CWCRT-eligible (dispatcher determined or officer) incidents/calls for service in 20 police areas among three of the cities participating in the CWCRT program, from January 1, 2022 through December 31, 2023. Our fixed-effects design strategy effectively compares outcomes conditional on effects associated with police area, year, month of year, day of week, and/or hour of day.

C. Sampling Plan

1. Existing Data

   Registration prior to analysis of the data

2. Explanation of existing data

   Regarding our total-effects research agenda, the data come from emergency dispatcher incidents/calls for service and criminal complaints and arrests provided by San Mateo County and/or police agencies in the cities of Belmont, Burlingame, Daly City, Foster City, Menlo Park, Pacifica, San Mateo, South San Francisco, and Redwood City. These data involving adults are available from January 1, 2019 through June 30, 2023.

   Regarding our isolated incident-level effects research agenda, the data come from emergency dispatcher incidents/calls for service and criminal offense and related arrest data provided by San Mateo County and/or police agencies in the CWCRT participating cities of San Mateo, South San Francisco, and Redwood City. These data involving adults are available from January 1, 2022 through December 31, 2023.

3. Data collection procedures

For both the total-effects and isolated incident-level effects research agendas, we received data transfers from San Mateo County or a particular focal city by means of signed data use agreements with the relevant governing bodies. We retain recorded emergency dispatcher call-for-service incidents and criminal complaints and arrests in each city police area in a given month over this time period.

4. Sample Size

Regarding our total-effects research agenda, our analytical sample consists of 48 distinct police areas in the cities of Belmont, Burlingame, Daly City, Foster City, Menlo Park, Pacifica, San Mateo, South San Francisco, and Redwood City, over 54 months, which is 2,592 police area-month observations from January 1, 2019 through June 30, 2023.

Regarding our isolated incident-level effects research agenda, our analytic sample consists of all CWCRT-eligible calls recorded by 911 dispatchers or police officers among 20 police areas in the cities of San Mateo, Redwood City, and South San Francisco over 24 months, which includes 4,118 CWCRT-eligible incidents/calls for service from January 1, 2022 through December 31, 2023—of which 1,208 involved an actual CWCRT co-response.

5. Sample size rationale

Regarding our total-effects research agenda, this sampling allows for observation of calls for service in the city 35 months before and 19 months after the beginning of the CWCRT program in a given police area, which allows for ample observation of pre and post treatment outcomes, tests of critical model assumptions, and for dynamic effects of the program.

Regarding our isolated incident-level effects research agenda, this sampling allows for observation of calls for service in the city 24 months after the beginning of the CWCRT program in three participating cities, offering a high-powered examination of all eligible calls recorded by 911 dispatchers or police officers for the duration of the program.

D. Variables

1. Measured variables

Regarding our total-effects research agenda, our Domain A1 outcome of interest is police area-month counts of CWCRT-related incidents (Chen and Roth, 2023). According to the CWCRT program's theory of change, the program targets to dispatch a co-response unit consisting of law enforcement officers and mental health clinicians to incidents that "involve a mental health component". This system relies on police officer and dispatcher discretion. This means we must necessarily code incident types according to whether an incident meets the definitions described in the theory of change.

Regarding our total-effects research agenda, the Domain A2 outcome is program-related criminal-offense counts, are instead differentiated by types of offenses that may be related to the program's stated goal of addressing situations involving a mental health-related crisis. We code criminal offense data using stated programmatic goals, which align with coding schemes from prior literature on these programs found in supplemental materials from Dee & Pyne, 2022. In response to these prior findings, we do not code assaults as program-related offenses, but include the original coding scheme with assaults included as program-related in our robustness checks. Using these data, we first construct a simple binary indicator equal to 1 for police area-month observations from police areas who participate in CWCRT during a given month (i.e., a "static" measure of treatment).

Regarding our total-effects research agenda, the Domain A3 outcome is program-related arrests, are also differentiated by types of offenses that may be related to the program's stated goal of addressing situations involving a mental health-related crisis. We code criminal offense data using stated programmatic goals, which align with coding schemes from prior literature on these programs found in supplemental materials from Dee & Pyne, 2022. In response to these prior findings, we do not code assaults as program-related offenses, but include the original coding scheme with assaults included as program-related in our robustness checks. Using these data, we first construct a simple binary indicator equal to 1 for police area-month observations from police areas who participate in CWCRT during a given month (i.e., a "static" measure of treatment).

Regarding our total-effects research agenda, the Domain A4 outcome is whether a 5150 emergency psychiatric hold was ordered for a given incident, which by definition constitutes a program-relevant incident.

Regarding our isolated incident-level effects research agenda, for the Domain B1 outcome we consider only criminal offenses recorded during an incident in which a dispatcher or police officer has deemed an incident CWCRT-eligible, regardless of whether the clinician was able to respond to the incident or not.

Regarding our isolated incident-level effects research agenda, for the Domain B2 outcome we consider only arrests made during an incident in which a dispatcher or police officer has deemed an incident CWCRT-eligible, regardless of whether the clinician was able to respond to the incident or not.

Regarding our isolated incident-level effects research agenda, for the Domain B3 outcome we consider only 5150 emergency psychiatric holds ordered during an incident in which a dispatcher or police officer has deemed an incident CWCRT-eligible, regardless of whether the clinician was able to respond to the incident or not.

E. Analysis Plan

1. Statistical models

Regarding our total-effects research agenda, our main confirmatory analysis in each of the four domains of the total-effects research plan is based on a difference-in-differences (DD) design, which assumes that CWCRT activity in a given police area and month leads to a constant, one-time change in the frequency of CWCRT-related calls for service for participating police areas (Roth et al., 2023; Chen and Roth, 2023). We do so by comparing changes in these outcomes among police areas participating in CWCRT to outcomes of police areas that either never participated or had yet to participate in CWCRT. The outcome will be a count of CWCRT-related calls for service. The predictors will be (1) an indicator of a treated police area in a treated month, (2) police area fixed effects, and (3) month fixed effects. Standard errors will be clustered at the police area level.

Although it is plausible that the program has a one-time static effect on our outcome domains of interest, we hypothesize a dynamic effect as the program becomes available. For example, this dynamic effect may increase related calls for service as knowledge of the program is more broadly recognized in pilot areas, and may then fall precipitously once the CWCRT team has been able to direct a critical mass of those in need to appropriate behavioral health services. For these reasons, we also use the timing of CWCRT participation to define less restrictive and flexibly dynamic measures of program participation. These include binary indicators for being one or more months after that first participation month. These measures flexibly allow for the initial participation in CWCRT to have effects that increase or decline over time.

Regarding the incident-level programmatic effects research agenda, our main confirmatory analysis is a fixed effects analysis differentiating program-eligible incidents receiving a CWCRT response versus those eligible but not receiving a response (i.e., due to a clinician being unavailable or off duty) by police area, time, day of the week, and hour of day (Gibbons et al., 2018).

2. Transformations

Regarding our total-effects research agenda, we use police area-month counts of program-related calls for service, counts of CWCRT-related criminal offenses, counts of arrests related to CWCRT-related criminal offenses, and counts of 5150 emergency psychiatric holds in the panel data to estimate the effects of the CWCRT program (Chen and Roth, 2023).

Regarding our isolated incident-level effects research agenda, we transform data so that we differentiate counts of criminal offenses, arrests and 5150 holds for all incidents recorded as program-eligible by 911 dispatchers or police officers involved in an incident. If the counts are non-zero and because we anticipate skewness in these data, we expect to transform the outcome variable into a natural log of the count for police area p in month m.

3. Inference criteria

   Regarding both the total and isolated incident-level effects research agendas, because we have a confirmatory outcome in each of the stated domains of interest, we will make inferences of our confirmatory analysis in each domain using two-tailed tests and p-values of $p<.10$. We will report p-values differently based on thresholds of $p<.01$, $p<.05$, and $p<.10$. If circumstances lead us to adopt multiple confirmatory outcomes in either of the domains anticipated, we will use the Romano-Wolf multiple hypothesis correction method for evaluating multiple comparisons in a particular domain.

4. Missing data

   Regarding the incident-level effects research agenda, we exclude data on one CWCRT participating city, Daly City, because in that city we only observe recorded reports of CWCRT eligibility if a clinician is available and responds to the call for service, and no information on eligible calls when a clinician is unavailable or off duty.

5. Exploratory analysis

   Regarding our total-effects research agenda, for each of the research domains in our total-effects designs, we will conduct a number of exploratory analyses. First, to test for time-varying treatment effects, we employ a semi-dynamic DD model that unrestrictedly allows for treatment effects unique to the month immediately after a police area first participates and up to 19 months later. We then test the equivalence of these coefficients of interest using the null hypothesis of a constant treatment effect.

   Second, we will conduct an "event study" analysis. A crucial maintained assumption of our DD approach is that the month-to-month outcome changes among comparison police areas (i.e., those without a change in treatment status) provide a valid counterfactual for what would have changed for treatment police areas in the absence of treatment. This "parallel trends" assumption is fundamentally untestable. However, we can provide qualified evidence on the validity of this important assumption through unrestrictive "event study" specifications that allow us to examine whether treatment and comparison group police areas had similar month-to-month changes in outcomes prior to the onset of treatment. To the extent that this hypothesis is true, it is consistent with the parallel trends assumption. We examine this question through the event-study specifications.

   Third, because these data also include counts of program-related incidents and criminal offenses that are unrelated to the CWCRT programs goals, there is an opportunity to test a "triple diff" (DDD) research design that allows us to account for unobserved disturbances in police area-month observations. Stacking our data at the police area-month (CWCRT & non-CWCRT) incident or offense level, the DDD specification includes fixed effects for all two-way interactions.

   Fourth, we analyze the confirmatory outcome during CWCRT-eligible and CWCRT-ineligible times.

Fifth, we explore a geography-based regression discontinuity design (e.g., Keele & Titiunik, 2015); several neighborhoods in treatment jurisdictions border comparison communities that do not yet offer co-response first-responder services. The proximity of geo-located 911 calls to these borders provides a sharp and discrete contrast in the availability of CWCRT services and an opportunity to estimate causal effects on outcomes. Relatedly, we can also improve the precision of this regression-discontinuity approach and address any possible other jurisdiction-specific confounds through leveraging a "placebo RD." Specifically, estimating the same regression-discontinuity designs with "pre" data allows for place-based fixed effects in "geographic difference in discontinuities" design (e.g., Butts, 2021).

Sixth, we explore the confirmatory outcome by available demographic characteristics (i.e., subject's race/ethnicity and gender).

Finally, in the event that crucial assumptions of the DD design are untenable, such as evidence that the parallel trends assumption may be violated, we will then consider other methods of evaluating program effects. This will include exploring comparative interrupted time series, counterfactual weighting, and synthetic difference-in-differences designs.

Regarding the incident-level programmatic effects research agenda, we explore the confirmatory outcome by available demographic characteristics (i.e., subject's race/ethnicity and gender).

F. Deviations from Pre-Registration Plan

At the request of CWCRT partners, we have included total effect estimates of the number of police cases opened among CWCRT-related calls in the total effects research agenda and among CWCRT-eligible calls in the incident-level effects research agenda.

Stanford | John W. Gardner Center for Youth and Their Communities

About the Gardner Center

Stanford's Gardner Center conducts research in partnership with school districts, nonprofits, foundations, and government agencies to generate practical solutions that advance equity for young people and their communities.

https://gardnercenter.stanford.edu

Create knowledge.
Ignite change.