

# More with Less: The Impact of Mandatory Overtime on Police Wellness and Productivity

Katie Bollman\* Ariel Gómez<sup>†</sup> Matthew Ross<sup>‡</sup> CarlyWill Sloan<sup>§</sup>

30 September 2024

[Please click here for the most recent version of this paper](#)

## Abstract

Can mandatory overtime provide an effective buffer against labor shortages? Or does the resulting work exhaustion prove counterproductive? We study a unique policy adopted by the Chicago Police Department (CPD) to evaluate the effect of mandatory overtime on officer well-being and productivity. The policy generated cross-sectional and temporal variation in overtime assignments, which we leverage in a difference-in-differences design. Using daily officer-level data, we find that, relative to the comparison group, treatment does not significantly increase officer injuries, absences due to sickness, or absence rates more generally. We also find little evidence that the policy had any effect on enforcement actions, including arrests, overall call response times, and use of force. Our findings are precise enough to rule out modest changes in several outcomes, including officer attendance, on-the-job injuries, and response times. For rare enforcement outcomes, our findings are less precise. Our results suggest that a limited use of mandatory overtime to manage labor shortages carries a low risk for officer health and productivity.

**Keywords:** Policing, Productivity, Public Safety

**JEL:** J21, J45, K42

## 1 Introduction

Staffing shortages and difficult work environments have strained public sector institutions across the United States, from schools and hospitals to police departments (Doan et al., 2024, Maryland Office of Legislative Audits, 2024, Pandemic Response Accountability Committee, 2023), concerns that have only intensified since the COVID pandemic. In law enforcement, vacancies in some departments—particularly large urban agencies—have fueled complaints about increased workloads for remaining officers and heightened concerns about burnout, morale, and performance (City of Chicago Office of Inspector

---

\*Oregon State University

<sup>†</sup>Northeastern University, School of Public Policy and Urban Affairs, Renaissance Park, 310, 1135 Tremont St, Boston, MA 02120, [ar.gomez@northeastern.edu](mailto:ar.gomez@northeastern.edu)

<sup>‡</sup>Northeastern University

<sup>§</sup>USMA at West Point

General, 2022, Police Executive Research Forum, 2023). These worries have grown amid shifting public sentiment and budgetary pressures (International Association of Chiefs of Police, 2020, National Conference of State Legislatures, 2021). However, while these perceptions and concerns are widespread, it is still unclear whether short-run increases in workload meaningfully impair frontline performance.

Most existing work focuses on the long-term consequences of stressful working conditions or on staffing outcomes like recruitment and retention. Evidence from other high-burnout occupations, mostly in the private sector, shows that cumulative fatigue and long work hours can impair worker productivity and increase turnover, adding to organizational costs (Blackburn et al., 2023), whereas extreme stressors have been shown to affect high-stakes decisions in law enforcement (Cho et al., 2023, Holz et al., 2023). . In practice, decisions by police leaders often involve modest, short-run adjustments — for example, canceling a scheduled day off — rather than wholesale changes to schedules or headcount. It is therefore plausible that small, temporary increases in workload may have limited impact, especially in agencies where officers are already accustomed to long hours. Understanding whether, and under what conditions, these marginal changes matter is crucial for evaluating how public-sector organizations can navigate perceived staffing shortfalls while maintaining public safety.

This paper directly addresses that research question: Do short-term, mandatory overtime assignments, implemented through cancellations of scheduled days off, actually influence officer behavior and performance on the job? We answer this question using administrative data from the Chicago Police Department (CPD), which, starting in 2022, implemented a new system of cyclical, pre-scheduled cancellations of Regular Days Off (RDOs). Under this policy, the first of officers’ two scheduled RDOs was canceled. As a result, affected officers worked five days in a row, followed by RDO, before resuming their regular schedule. These cancellations were introduced to address chronic staffing shortages and were applied broadly across bureaus and assignments according to a predetermined calendar. While the CPD has long used separate, event-driven RDO cancellations in response to large gatherings or anticipated public safety threats, the cyclical system we study was distinct: it was not tied to specific events, determined in advance, and affected thousands of officers at once. As a result, it generated plausibly exogenous variation in officer workloads, unrelated to individual officer behavior or short-term operational needs.

To estimate the effects of these cancellations, we use linked personnel, dispatch, and scheduling data from 2021 to 2023. We identify RDO cancellation events using department-wide schedules and construct a balanced panel of officers with rotating work-rest patterns. Our empirical strategy considers officers with otherwise identical work schedules, comparing officers who have one RDO cancelled to others assigned to the usual work schedule. We compare the outcomes of officers in the first group to those in the second in the days following an RDO cancellation. Our specification implements a stacked difference-in-differences design with officer, date, and assignment fixed-effects. This approach allows us to isolate the causal effect of short-term, externally

imposed increases in workload on officer behavior. This setting offers several advantages. First, it isolates short-run, plausibly exogenous variation in workload, avoiding common confounds related to officer selection into overtime or high-activity shifts. Second, it allows us to track outcomes with high temporal precision. Third, it provides insight into how police organizations adapt to chronic staffing shortages in high-stakes environments.

We find that being assigned to an RDO cancellation significantly increase the likelihood that officers actually work on an RDO, confirming broad compliance with the policy. However, these cancellations do not appear to affect officer behavior or productivity across key outcomes. In the work week following an RDO cancellation, officers respond to the same number of calls for service and arrive just as quickly; both of these effects are precise zeros. We also find no change in discretionary enforcement activity, including stops, arrests, and use of force, though these estimates are less precise. Heterogeneity tests reveal no significant differences in treatment effects between officers by demographic characteristics, enforcement patterns, or the number of times they are assigned treatment. Results are robust to a range of modeling choices and we find no evidence of pre-trends or effects on placebo dates. XX would like to add to this paragraph what we rule out XX

This paper contributes to several strands of literature on overtime, burnout, and police performance. In health care and other frontline occupations, research has documented that long hours and stressful working conditions reduce productivity, impair judgment, and can negatively impact health and safety. Studies such as Blackburn et al. (2023), Nekoei et al. (2024), Sachiko and Isamu (2016) show that sustained overtime can degrade performance and increase errors. Early theoretical work also suggests that when burnout is a concern, organizations should limit overtime and increase hiring to maintain output quality (Ehrenberg, 1970, Yaniv, 1995). Together, this research implies that policies relying heavily on mandatory overtime as a substitute for adequate staffing may produce diminishing returns.

Despite these broad findings, relatively little work explores these questions in the context of public-sector employees or police officers. Existing evidence highlights the unique constraints faced by public-sector agencies like police departments, which often have long-term hiring processes and limited flexibility in recruiting replacements (Linos, Ruffini, and Wilcoxon, Linos et al., Maryland Office of Legislative Audits, 2024, San José Office of the City Auditor, 2016, Stoughton, 2016, Whitaker, 2023). Because law enforcement involves high-stakes decisions and serious risks to public safety, the consequences of officer exhaustion may also differ. Descriptive evidence suggests that fatigued officers may be more prone to errors and biased decision-making, and more likely to incur injuries on the job (James, 2018), which may in turn reduce public safety and police legitimacy (Cho et al., 2023, DeAngelo et al., 2023, Holz et al., 2023). Experimental evidence also indicates that police performance varies under different scheduling structures, and that perceptions of fairness and leisure matter for labor supply decisions (???). This underscores the need for careful identification of causal effects of overtime on officer behavior.

Our project is most closely related to Ferrazares (2025), who examines the cumulative effects of consecutive workdays on police behavior using CPD shift data. He finds that each additional day worked increases the risk of use of force and officer injury, while reducing enforcement activity, patterns consistent with fatigue and declining capacity. Our design differs in both identification and focus. Rather than tracking outcomes across consecutive days within a regular workweek, we isolate the short-term causal impact of involuntary overtime by leveraging pre-scheduled RDO cancellations. These cancellations, set independently of officer behavior or preferences, provide plausibly exogenous variation in workload. Importantly, we study outcomes in the days following an RDO cancellation, not on the day of the cancellation itself, which is always followed by a non-cancelled RDO before officers return to duty. Our precise null effects do not contradict the findings in Ferrazares (2025): if fatigue accumulates over multiple consecutive days, then a policy in which a single extension is followed by a rest day may avoid such consequences. Our results suggest that ensuring adequate recovery time after short-term overtime may mitigate the behavioral costs of increased workload. By leveraging this rotating assignment of mandatory overtime across different officer groups and units, our study offers new evidence on the consequences of overtime in public safety settings and can inform policy decisions facing other departments with similar labor constraints.

Our findings have practical implications for police leaders navigating persistent staffing challenges. While public discourse often focuses on changes in recruitment or large-scale reforms, most staffing decisions involve more limited, incremental changes. In practice, labor shortages are typically managed at the margin through overtime and extended shifts, rather than restructuring the work week or increasing staffing. The CPD’s systematic and cyclical RDO cancellations offer an example of such marginal adjustments without the confounds of immediate operational demands. Understanding the effects of these types of changes helps clarify what can realistically be achieved within existing operational constraints.

Our setting closely mirrors these realities. In 2021, the year before the day-off cancellation policy was put in place, Chicago police officers already averaged 11 hours of overtime on top of the usual 54 hours per week, and nearly 10% of officers worked more than 16 hours of overtime per week in that same year. Given this already-high baseline, it would likely take a substantial change in workload before observable effects on performance or public safety emerge. Our results suggest that modest increases in mandatory overtime do not produce measurable short-run harms, just as small reductions likely would not yield meaningful improvements. Sustained or structural changes may still affect burnout or morale, but for the kinds of incremental policy adjustments most departments face, our evidence indicates that small shifts in overtime use are unlikely to materially worsen or improve outcomes. More broadly, these findings may extend to other public-sector workforces where overtime is a common stopgap and where large-scale reforms remain difficult to implement.

## 2 Context

Police departments across the U.S. have experienced persistent labor shortages in recent decades. More recent data show a steeper decline since 2020; between 2019 and 2022, agencies reported a nearly 50% increase in resignations and a 20% increase in retirements among sworn police officers. Combined with challenges in recruitment, these declines led to a reduction of 5% in total staffing of sworn officers between 2020 and 2023 (Police Executive Research Forum, 2023).

The CPD, which polices the third largest U.S. city (U.S. Census Bureau, 2024), has been especially affected by these staffing declines. Between 2019 and 2022, total retirements and resignations among sworn Chicago police officers nearly doubled. Over the same period, total staffing of patrol officers declined by more than 10%. Although it is difficult to cleanly identify the specific causes of these declines, they coincided with a social upheaval that accompanied both the COVID-19 pandemic and a wave of protests against excessive use of force by police officers.

In May 2020, facing significant police vacancies and large protests against the killing of George Floyd, the CPD began canceling previously scheduled days off to ensure there was sufficient staffing during surges in crime and disorder. Over the subsequent three years, the CPD repeatedly mandated police overtime through these cancellations and instituted temporary increases in working hours, leading to complaints by staff. In an audit of the CPD’s employee records, the Chicago Office of Inspector General reported that some officers had worked as many as 12 consecutive days with shifts as long as 12 hours in April and May of 2022 (City of Chicago Office of Inspector General, 2022).

Beginning in 2022, the CPD responded to public criticism over the arbitrary mandatory overtime assignments by shifting to a policy with pre-determined day off cancellations that would cycle through groups of officers.<sup>1</sup> This policy was in place until October of the following year, when newly-appointed Superintendent Larry Snelling announced a general prohibition of mandatory overtime except for holidays and special events (Sisk, 2023).

To examine the relationship between mandatory overtime and officer-level outcomes, we leverage this policy and its cyclical assignment pattern in a stacked difference-in-differences design. Understanding our empirical strategy requires some institutional background on how CPD officer schedules are determined and how mandatory overtime was allocated throughout our study period.

To ensure that enough officers are on call every day of the year, the CPD assigns all officers to groups with schedules set annually. These “day-off groups” (DOGs) rotate between working and off days, with each DOG scheduled to work four consecutive days

---

<sup>1</sup>Mandatory overtime triggered by special events or during certain holidays were decided independently of this policy.

followed by two consecutive regular days off (RDOs).<sup>2</sup> We study officers assigned to six DOGs (numbers 61-66) whose schedules are sequentially offset by one day. For example, when officers in DOG 61 are scheduled for their second consecutive day off, officers in DOG 62 will be scheduled for their first day off, and DOGs 63-66 will be assigned to report for duty. The schedule rotates such that the following day, groups 61 and 64-66 are scheduled to report for duty while 62 and 63 are scheduled for rest. DOG assignments are determined through a seniority-based bidding process in the fall of the previous year. While officers can and do change DOGs throughout the year – we observe switching in nearly 40% of officer-years– they do not have direct control over new work schedules, which are up to the discretion of management.

Officers are also given both a geographic and unit assignment. The CPD divides the city into three main administrative levels. The largest level is an area, such that the city of Chicago is split into five distinct patrol areas. Areas are then divided into 22 districts, which are further subdivided into beats. Unit assignments may entail patrolling within a specific district, or to non-patrol tasks that do not adhere to district boundaries. New recruits are assigned to a given district or unit and can eventually bid for assignment changes based on seniority.

During our study period, RDO cancellations were assigned to officers in accordance with “cancellation matrix” calendars, which assigned mandatory overtime in a set, rotating pattern. For a given cancellation cycle, a subset of areas would be selected, and within those areas, cancellations would cycle through each DOG. The next cycle would shift to a different subset of areas and cycle through each DOG and so on. This rotating pattern of RDO cancellations induces plausibly exogenous variation in the type of officer working on a day off.

### 3 Data

Our empirical analysis relies on several administrative datasets, which we obtain through Freedom of Information Act (FOIA) requests, combined with public data sources. Our main administrative data comes from the Attendance and Assignment (A&A) database, in which each observation represents a single officer on a given date. These data also include information on shift (watch), unit and beat assignments, whether an officer reported for work, and the reason for any absence. To determine DOG assignments, we use another administrative database that lists the dates each officer was assigned to a given DOG. RDO cancellation dates come from “cancellation matrices” obtained from the CPD, which show how cancellations cycle through area and DOG combinations. To determine treatment assignment for each officer on a given date, we combine these data with information extracted from operations calendars, which determine work schedules for each DOG each year.

---

<sup>2</sup>Recent work has also leveraged the CPD’s DOG system to study the roles of officer race, gender and experience in determining officer performance (Ba et al., 2021, Gudgeon et al., 2024, ?).

In A&A data from 2022 to 2023, some officers who report for duty are given an absence code corresponding to an RDO cancellation, allowing us to check for compliance with treatment. We supplement attendance data with information on overtime, which also indicates whether overtime was voluntary or mandatory. To assess outcomes, we combine attendance data with officer-date-level data on enforcement activity, including arrests and 911 calls, also obtained from the CPD.

We restrict our sample to observations recorded between April 1st, 2022 and June 30th, 2023, the last date for which A&A data is currently available.<sup>3</sup> We further restrict our analysis to six DOGs that comprise 66% of patrol officer-date observations in our sample. This conditioning simplifies the analysis, as each of these DOGs follows a consistent four-days-on, two-days-off pattern.<sup>4</sup> Finally, we also restrict our sample to units assigned to the five main geographic areas, Areas 1 through 5. Units assigned to other “areas” perform city-wide tasks, such that their duties may take place in multiple areas in a given day.

Table 1 presents summary statistics for our short stacked sample. We observe 7,566 distinct officers between April 2022 and June 2023, adding up to more than 2.4 million total observations. 54% of the sample reports to work on an average day, whereas two percent of officer-days are reported to include involuntary overtime hours. Unscheduled absences are not uncommon: 3.6% of the sample is absent due to injuries on the job while officers call in sick about 5.8% of the time. 8% of officer-dates are marked absent due to officers taking paid time off. On any given day, about .02% of the sample reports a different DOG assignment from the day before, whereas another .02% report a change in area assignment. For enforcement actions, some outcomes are more common than others. Across all observations in our sample, including days when officers are absent, officers report close to 1.9 responses to calls for service on an average day. For total investigatory stops (similar to stop-and-frisk) and total arrests, the respective means are about .04 and .07. The data show that, even when they report to work, officers do not arrest or stop anyone on most days. The probability that an officer employs force across the entire sample is about .1%.

To supplement the A&A data, we also use an administrative dataset that compiles all reported overtime shifts for each officer. In the following figures, we present descriptive plots to offer context on how overtime use has changed since 2014 and during our study period. These figures only include officers who have worked at least one overtime shift in the specified period, the relevant population in our analysis.

In Figure 1a, we plot the average number of overtime days worked per officer, which increased by about 40% between 2014 and 2023. Breaking down overtime by type in Figure 1b, we can see that while there is a consistent increase in involuntary overtime

<sup>3</sup>Although we observe cancellation matrices for February and March 2022, some of the assignment patterns for these months are not clear. Moreover, the data are not fully consistent with either the explicit or implied cancellation patterns until April 1st, 2022.

<sup>4</sup>In comparison, the work cycle for DOGs 71 to 77 alternates between four days working and three days off and five days working with two days off in a pattern that repeats every seven weeks. The other major set of DOGs, 8 to 10, do follow a regular schedule, but get the same weekdays off every weekend. By focusing on the 60s DOGs, we ensure that any effects are not weekend-specific.

through 2020, it is actually voluntary overtime that increases the most since then.<sup>5</sup> Breaking down overtime days per officer at the monthly level during our study period, from 2022–2023, ?? and Figure C.18 display notable patterns that help to contextualize the policy environment during our study period. Consistent with the other figures, total overtime and voluntary overtime are gradually increasing each month. Involuntary overtime is mostly flat, but shows a large spike in the summer months of 2022 and a much smaller bump during the summer of 2023, both of which can partially attributed to seasonal fluctuations. But the sharp increase in 2022 is consistent with reports about excessive overtime and the resulting public backlash against CPD policy.

## 4 Empirical Strategy

Our empirical design relies on the quasi-random assignment of officers to overtime introduced by RDO cancellations. The cyclical pattern of these cancellations, combined with DOG and area assignments for each officer, generates within-DOG and between-area variation. We consider each RDO cancellation to be a single natural experiment, or “sub-experiment”. By pooling data across sub-experiments in a stacked sample, our stacked difference-in-differences design allows us to leverage all RDO cancellations we observe to estimate the effects of mandatory overtime on a range of officer-level outcomes. We discuss the construction of our stacked sample in the following sub-section.

### 4.1 Stacked Sample Construction

In our sample, RDO cancellations always fall on the first RDO of a given DOG-area pair. Complying officers report for mandatory overtime on this date and have the following day off before beginning the work cycle anew. For example, on May 31st, 2022, the CPD cancelled the RDO for officers assigned to DOG 61 and Area 1. In our stacked difference-in-differences design, the set of officers we observe on this date who are assigned to DOG 61 constitute a “sub-experiment”: officers from DOG 61 assigned to Area 1 are treated, whereas officers from DOG 61 in all other areas follow their usual work schedules and serve as counterfactuals.

To construct a trimmed stacked sample pooling data across sub-experiments, we follow Wing et al. (2023). For each RDO cancellation date, we take a subsample of all officers in the treated DOG and define a treatment indicator based on whether they were assigned to the affected area. Through this approach, we create two stacked samples. In the short-run sample, we consider data from the five days preceding treatment to the five days following the RDO cancellation. In a longer-run sample, we include 12 days preceding and 11 days following a given treatment date, such that two full work cycles are included in the each of the pre- and post-treatment periods. In either case, we center the data in each sub-experiment with respect to a given cancellation date, such that

---

<sup>5</sup>Data breaking down overtime shifts by voluntary and involuntary is not available until 2018. Because the policy was only in place from 2022–2023, data on Tiered Deployment is only available in those two years.



each day is defined relative to treatment. For instance, the day of the cancellation will be set to zero, whereas the day before is set to -1 and the day after to 1. Data from each of these sub-experiments are then stacked to create a single stacked dataset. To ensure that our results are not driven by compositional changes in the officers observed on each day of a given sub-experiment, our sample only includes officers that are observed on all days of said sub-experiment.<sup>6</sup>

Table 1 presents summary statistics for our analytical sample. We observe 5,606 distinct officers over eleven days across 215 sub-experiments, adding up to more than 1.26 million total observations. Thirty-eight percent of the sample is assigned to treatment and 54% report to work on an average day. Two percent of officer-days are reported to include involuntary overtime hours, most of which occur when officers work on the date of the cancelled RDO that defines the sub-experiment. Unscheduled absences are not uncommon: 4.5% of the sample is absent due to injuries on the job while officers call in sick about 6.5% of the time. 9.4% of officer-dates are marked absent due to officers taking paid time off. On any given day, about 2% of the sample reports a different DOG assignment from the day before, whereas around 1% report a change in area assignment. For enforcement actions, some outcomes are more common than others. Across all observations in our sample, including days when officers are absent, officers report close to 2.5 responses to calls for service on an average day. For total investigatory stops (similar to stop-and-frisk) and total arrests, the respective means are about .06 and .03. The data show that, even when they report to work, officers do not arrest or stop anyone on most days. The probability that an officer employs force across the entire sample is about .16%.

## 4.2 Empirical Specification

With our stacked dataset, we adopt a saturated difference-in-differences specification, as recommended by Wing et al. (2023):

$$Y_{ier} = \beta_0 + \beta_1 D_{ie} + \sum_{h \in [-5, 5] \setminus \{-1\}} [\gamma_e \mathbb{1}\{e = h\} + \delta_e D_{ie} \times \mathbb{1}\{e = h\}] + \epsilon_{ier} \quad (1)$$

Here, an outcome  $Y_{ier}$  for officer  $i$  in sub-experiment  $e$  at event time  $r$  is regressed on a set of indicators for treatment assignment,  $D_{ie}$ , days relative to the given cancellation,  $\mathbb{1}[e = h]$ , and their interactions. These regressions exclude indicators for the day before a cancellation, such that the parameters are estimated relative to the difference in treatment and control at event time -1. The set of  $\delta_e$  on the interactions of treatment and event time are our coefficients of interest. To reduce concerns over endogenous assignment changes, we impute area and DOG data with an officers modal assignment

---

<sup>6</sup>Recall that we can observe working and absent officers. For example, if an officer calls out sick on a work day, we observe an absence code indicating the reason for their absence. When there is a gap in an officer's panel, either their absence was not recorded or we were unable to fully link their information across data sources.

in each month. We consider officers who change assignments within a given month to be non-compliers, and therefore interpret our coefficients of interest as intention-to-treat (ITT) estimates.

Wing et al. (2023) show that because observations are reused in different sub-experiments, unweighted regressions implicitly put different weights on treatment and control observations. We therefore follow their recommendations and use sample share weights to ensure that treatment and control observations are equally weighted in the regression. Treatment observation weights 2 and control observation weights 3 are defined as follows:

$$\frac{(N_e^D + N_e^C)/(N^D + N^C)}{N_e^D/N^D} \quad (2)$$

$$\frac{(N_e^D + N_e^C)/(N^D + N^C)}{N_e^C/N^C} \quad (3)$$

$N_e$  is the number of total observations in sub-experiment  $e$  and  $N$  is the size of the total stacked sample.  $N^D$  and  $N^C$  represent the number of observations in the treatment and control groups, respectively. With these weights, our main specification will identify a weighted average of ITT estimates for each  $\delta_e$ , shown in 4. This specification effectively aggregates the sub-experiment ITT estimates with weights determined by the share of observations that each sub-experiment contributes to the overall sample.

$$\sum_{\Omega_e} ITT(e, r) \frac{N_e^D + N_e^C}{N^D + N^C} \quad (4)$$

Each sub-experiment compares observations within a DOG, so treatment is assigned at the area-level. Because there are only five units, however, analytical clustered standard errors are unlikely to provide valid inferences. We therefore estimate area-clustered standard errors using a wild bootstrap.

### 4.3 Threats to Identification

To identify the causal effects of overtime on officer productivity, our empirical strategy must account for two main sources of endogeneity. First, because overtime is normally voluntary, individuals who work extra hours are likely to be systematically different from those who do not. By studying *mandatory* overtime, we can account for this selection. Moreover, because we estimate intention-to-treat effects, selection resulting from imperfect compliance will not bias our estimates. Second, in the context of law enforcement, the demand for overtime is likely to be correlated with spikes in crime.

Then differences in outcomes between days when an officer works more versus fewer overtime hours are likely to be confounded by trends in criminal activity. In our context, because RDO cancellations occur in a sequential pattern throughout the year, the confounding effects of date-specific variation in crime is likely to be attenuated.

By comparing individuals across areas, our empirical design raises an additional selection problem: because they experience different working conditions, officers in different areas of Chicago are not directly comparable. For example, crime rates likely vary across areas, which could in turn affect relevant outcomes, such as on-the-job injuries. Furthermore, area assignments are not random, such that officers may differentially select into areas based on observed and unobserved characteristics. Through our daily officer panel, however, we can account for initial level differences in officer assignments and characteristics in a difference-in-differences design.

Although our identification strategy can account selection stemming from time trends and initial differences within each sub-experiment, it is only valid under three assumptions. First, there must be no anticipation effects, meaning that officers cannot change their behavior in anticipation of a scheduled RDO cancellation. There are at least three reasons this assumption may be violated. One possibility is that officers may take unscheduled absences in advance of a cancellation to avoid treatment, for example, by calling in sick or taking an unexcused absence. Our results, however, will show that there is no evidence of anticipation effects for absences.

A second possibility is that officers choose their schedules to evade treatment or to avoid mandatory overtime on specific days, including holidays and dates with higher crime rates. The CPD’s process of assigning yearly schedules, however, makes this possibility unlikely. Each November, officer work assignments for the following year are determined through a seniority-based bidding system. Although more senior officers may have some choice over the dates on which they must report for work, no officer has full discretion over their schedule, which is mostly determined by the day-off pattern of their assigned DOG. For example, an officer who selects a DOG with RDOs falling on December 24th and 25th cannot also have an RDO land on New Years Day. Moreover, even with limited discretion, it is unlikely that officers fully anticipate changes in crime (Ba et al., 2021) and the full set of RDO cancellation matrices for the coming year when bidding for schedules. Therefore, conditional on officer assignments to DOG and unit, whether an officer is scheduled to work on a given day is likely orthogonal to conditions on the ground in the areas they are assigned.

Because official documents from the CPD state that officers can petition to change work assignments throughout the year, officers may also evade an upcoming RDO cancellation by switching to a new area or DOG. For example, about 17% of officers in 2022 are assigned a different DOG on December 31st than on January 31st. An important caveat to these changes is that new assignments for other job characteristics, such as vacation and furlough dates, are at the discretion of unit managers. If officers receive assignments close to their preferences under the bidding system, the individual costs of changing DOG or unit assignments may be higher than the benefit of avoiding

an RDO cancellation. Moreover, since mandatory overtime rotates through each DOG-area pair, an officer that changes assignments to avoid a cancellation will eventually experience treatment anyway. Even so, we will show that treatment officers do not differentially change work assignments ahead of an RDO cancellation.

In addition to no anticipation effects, our empirical strategy assumes that differences in outcomes between treatment and control officers would have remained constant following RDO cancellation dates in the absence of the policy. Although we cannot directly test this assumption, if outcomes do not evolve in parallel ahead of a cancellation date, then the parallel trends assumption would be unlikely to hold. Conversely, because we find that treatment and control officers follow comparable pre-treatment trends for all outcomes, this assumption is plausible. Even if outcomes follow common trends, an RDO cancellation may coincide with an unexpected event, such as a localized spike in crime, that puts either treatment or control on a different path. Because cancellations follow sequentially and are spread throughout the year, however, we expect the effects of such shocks to attenuate when we aggregate across sub-experiments.

Identification in our setting requires an additional assumption restricting the dynamics of treatment effects. Because all DOG-unit pairs are eventually assigned to treatment, if a single RDO cancellation has a permanent effect on outcomes, comparison groups in later sub-experiments would be contaminated by previous treatments, invalidating our empirical design. ITT effects can therefore only be identified if treatment effects are short-lived. Our main specification therefore assumes that treatment effects dissipate after a full work cycle, or six days. Under this restriction, the full control group of a given sub-experiment may be composed of both contaminated and uncontaminated control observations. Our sample will therefore drop officers assigned to DOG-area pairs that were treated in the six days prior to the RDO cancellation in each sub-experiment. If we impose the same restriction on our longer-run sample, we can include control observations that were not assigned to treatment within thirteen days of a sub-experiment's cancellation date

We argue these restrictions are plausible for two reasons. First, officers are assigned at least two consecutive days of rest before they are considered valid controls, potentially limiting the persistence of treatment effects. Second, barring changes in schedule or geographic assignments, officers will be assigned mandatory overtime no more than once a month. As a result, the cumulative effects of consecutive treatments are likely to wash out as officers experience other stressors and periods of rest in between each day off cancellation. Though these restrictions cannot be directly tested, we can loosen our restriction in our short-run specification, such that treatment effects dissipate after two full work cycles, or twelve days. We will show that our short-run results are comparable regardless of which restriction is imposed.

## 5 Results

### 5.1 Binscatter Plots

Before presenting our formal regression results, we first present binned scatterplots of the raw stacked data. In the following figures, we plot the average level of a given outcome by treatment and control officers for each day relative to the RDO cancellation date in each sub-experiment. In Figure 2a, we plot the proportion of officers in the stacked sample reporting for work on each day relative to an RDO cancellation separately for treatment and control. The pattern in attendance is consistent with the officer work cycle as described in official CPD documents: officers are normally scheduled to work four days in a row and take two consecutive days off before restarting the work cycle. Based on information from CPD operations calendars, we should expect relative days -5, 0, and 1 to be regular days off, whereas all other days are regular work days. Officers in the control group clearly follow this pattern, with rates of attendance being about 50% lower on RDO days.

On the day of the RDO cancellation, however, we can see that about half of officers in the treatment group report to work, about 30% more in compared to officers in the control group. We interpret this as evidence of partial compliance with the RDO cancellation policy. Figure 3a shows a similar pattern in overtime hours. By definition, any hours worked on an RDO date will be counted as overtime, even if officers are mandated to report to work. On the cancellation date, treated officers report almost four overtime hours on average, whereas control officers report about one and a half hours on average. Given that a regular work day for the officers in our sample lasts a total of nine hours, if about half of officers in the treatment group report for mandatory overtime on day 0, as suggested in Figure 2a, we should expect about four and a half overtime hours on this day for the treatment group. The discrepancy may be partially explained by the fact that, in both treatment and control groups, some officers report voluntary overtime work. Then it is likely that not all officers who report for work on the RDO cancellation date are working mandatory overtime. Nevertheless, in results not reported here but available upon request, about 35% of treatment officers report working any involuntary overtime hours on the treatment date. Because not all officers assigned to treatment comply with the policy, we consider any results as intention-to-treat (ITT) estimates.

In addition to treatment compliance rates, the plots in ?? also reveal that, on all days other than the cancellation date, officers in the treatment group report nearly identical attendance rates. This finding suggests, on one hand, that treatment and control officers are on comparable, parallel paths, and on the other, that RDO cancellations do not have a short-run effect on the probability of working. Officers do not appear to either skip work in anticipation of an RDO cancellation nor do they make up for the lost day by skipping work in the following work cycle. The raw data are therefore consistent with the parallel trends and no anticipation assumptions and suggests that officers do not undo the effect of the policy by taking personal days in the short-run.

Officers may additionally evade treatment by changing their work assignments, switching to either a different area or DOG. In ??, however, it can be seen that overall assignment changes are rare from one day to the next. Moreover, trends for both treatment and control roughly follow each other and there are no clear trend breaks around the treatment date.

In the next few figures, we report more binned scatterplots for some of our main outcomes. In ??, we see that treatment and control officers follow similar trends in illness-related absences both before and after treatment, suggesting that, at least in the raw data, there is little short-run evidence of increased illness among officers in response to an RDO cancellation. We next consider enforcement outcomes in ?. First note that arrests are sufficiently rare that, on a regular work day, officers are only about 4% likely to make a single arrest. Second, officers deal with calls for service on a more regular basis, responding to three calls on an average regular work day. In both cases, we see that treatment and control track each other closely, suggesting there may not be an effect of RDO cancellations on enforcement. Finally, ? reports binned scatter plots for the average number of seconds it takes for an officer to arrive at the scene of a call for service. Again, treatment and control officers appear to track each other, suggesting there is no effect on response times.

Given how treatment cycles across groups of officers, identification in our setting relies on a restriction on the dynamics of treatment effects. The preceding plots include only officers who were last assigned treatment at least six days before day -5. Under the assumption that treatment effects dissipate after six days, the control group in the binned scatterplots would not be contaminated by the effects of previous RDO cancellations. Although this assumption cannot be directly treated, we can instead exclude officers who were treated within 12 days of relative day -5, as we do in APPENDIX SECTION. In this subsample, control officers may be considered clean counterfactuals under the assumption that treatment effects dissipate after 12 days. Under these conditions, we again see similar patterns as before: clear evidence of partial compliance with the policy and comparable outcome trends between treatment and control officers for all days besides the RDO cancellation date.

Although data limitations do not allow us to restrict the control group further, it is not unreasonable to think that the effects of a single day of mandatory overtime do not have lasting effects beyond two complete work cycles. Moreover, it is difficult to construct a story of persistent treatment effects beyond 12 days that can explain roughly parallel trends before and after treatment. One possibility is that the effect of treatment is an absorbing state, such that, once treated, an officer experiences a permanent change in outcomes, with no further effects of additional RDO cancellations. Because most observations in our stacked sample are post-initial treatment, under this assumption, we should expect parallel trends in outcomes around future RDO cancellation dates. If it were true, however, that treatment effects were permanent, one would expect that they would be cumulative however, in which case it would again be difficult to explain the patterns we see in the raw data.

One other concern with the patterns observed so far is that five days before and after treatment may be too short of a timeline to observe any effects. For example, if treated officers wish to take a personal day to make up for their lost day off, they may not call in sick in the five days after being assigned mandatory overtime, as their supervisors may suspect that they are trying to make up for lost leisure time. Sophisticated officers may therefore wait until the following work cycle to make up their cancelled day off by calling in sick. Similarly, if officers take a personal day in anticipation of an upcoming RDO cancellation, they may prefer to take it before the work cycle preceding the cancellation to not arouse suspicion.

We address this concern with additional figures in the APPENDIX, in which the event-study plot is extended to include two full work cycles preceding and following an RDO cancellation. The control officers included in these plots are at least six days out from their last treatment date from relative day -12 and are clean counterfactuals under the assumption that treatment effects dissipate after a full work cycle. Even with this extended timeline, we see similar patterns as in the previous figures, with roughly parallel trends before and after the treatment date, suggesting there are not anticipatory or post-period effects on our main outcomes.

## 5.2 Formal Regression Results

In this section, we assess the effects of involuntary overtime on officer outcomes through formal difference-in-differences specifications. In the following event study plots, each point represents the coefficients on the interaction between treatment and event time indicators from a dynamic specification like Equation 1. The value of each estimate gives the average difference in the outcome between treatment and control for a given event-time day, relative to the difference on day -1, which we normalize to zero. The salmon-colored line delineates the pre-period from the treatment day while the error bars report the 95% confidence intervals estimated from a wild boot strap. For comparison, I also plot in blue the estimate from a simple difference-in-differences specification, interacting treatment with a post-period dummy.  $\beta$  gives the value of the coefficient and  $\mu$  gives the mean of the given outcome in the pre-period.

## 5.3 Treatment Compliance

We begin by estimating the “first stage” effect of being assigned to treatment on actually working on a canceled RDO. ?? includes data from five days before treatment and only one day in the post-period, the day of treatment itself, in order to more cleanly estimate treatment compliance. ?? shows that, relative to event time -1, officers are about 30 percentage points more likely to report to work when they are assigned to treatment. Although the 95% confidence interval on the estimated effect is wide, the effect is clearly statistically different from zero. The simple difference-in-differences estimate, shown in blue on the plot, reports a large t-statistic, just below seven. The estimated effect is also large relative to the pre-period mean of about .57, which includes both regular

work days and an RDO date. The lower bound on the estimate on the day of treatment is just over .1, implying an effect about 20% of the pre-period mean.

The coefficients for the days leading up to the treatments date are small and close to zero in comparison. Moreover, given how tight the confidence intervals are, especially for regular work days, we argue that pre-period differences are precisely estimated zeros. These results are consistent with parallel pre-trends, suggesting that treatment and control officers are comparable and supporting the plausibility of the parallel trends and no anticipation assumptions. In ??, we plot the coefficients for the same specification, now with total overtime hours on the left-hand side. The results in this plot are consistent with ?? and provide further evidence supporting our empirical design. Additional plots in the APPENDIX restricting the sample to officers that have not been treated for at least two work cycles and extending the timeline of the plot to two work cycles before treatment show similar results.

## 5.4 Effect on Officer Outcomes

We next assess the effects of RDO cancellations on officer outcomes after treatment. In the following event study plots, we now include coefficients for the effects on the days following day 0 but exclude coefficients for the actual day of treatment. As shown in ??, treatment and control officers report to work at different rates on at event time 0. Comparisons in outcomes that can only be observed when officers are at work will therefore be invalid. Thus, our analysis on the effects of mandatory overtime will focus on outcomes observed after event time 0.

In ??, the event study plots show that there are no statistically significant effects on either reporting to work or in the number of overtime hours worked by officers in the five days after treatment. The confidence intervals are also modest relative to the pre-period sample means. For example, in Figure 2b, a decrease in the probability of officers reporting to work larger than .04 times the pre-period mean can be ruled out with 95% confidence. The estimated effects on overtime hours are less precise but can still rule out large positive effects. We report simple difference-in-differences specifications for these outcomes in Columns 1 and 2 in Table 3 and also find null results with narrow confidence intervals.

Given that we do not find significant changes in reported absences in response to RDO cancellations, one might not expect to find effects on specific categories of absences. Nevertheless, in ??, we report event study plots for illness-related absences. As expected, we do not find overall significant effects on either on-the-job injuries or calling in sick. The estimates for calling in sick are much less precise relative to the estimates in ??, but we can rule out modest effects on injury-related absences. Regression results from Columns 3 and 4 in Table 3 are consistent with our event study plots. In Column 5, we also report effects on the probability that officers take paid time off and find similar results.



Thus far, our estimated effects suggest that the CPD’s mandatory overtime policy was successfully implemented without causing significant harm to the average officer assigned to treatment. We do not find any evidence that officers are undoing the effect of the policy by taking additional days off either before or after an RDO cancellation, at least in the short run. In the APPENDIX, we report corresponding event study plots and regression tables for the restricted and longer-run samples and find similar results, though we lose precision in our estimates when we use the longer sample. Even taking into account that compliance rates on the day of treatment are only about 30%, the implied local average treatment effect on reporting to work based on the point estimate in Table 3 is only about a 2% decline relative to the pre-period mean. We do not push this interpretation given that the exclusion restriction is unlikely to hold, but this back-of-the-envelope calculation further underscores the magnitude of our estimates. Although we of course cannot completely rule out the possibility that officers were injured as a direct result of this policy, the results suggest that officers do not generally experience above-average physical harm after an RDO cancellation.

Even if the overtime policy does not reduce total officer-work days or harm officers at rates above the average, officers may still experience fatigue as a result of an RDO cancellation, possibly affecting their productivity and ability to effectively respond to threats to public safety. In ??, we report event study plots estimating the impacts of treatment on the likelihood that an officer performs either of the two most common law enforcement actions: responding to 911 calls and making arrests. We do not find evidence of statistically significant effects in the post-period for either outcome. Moreover, trends look about equally as flat in the pre-period as in the post-period for both outcomes. We do find important differences in the precision of the estimates. The effects on responses to calls for service are narrow relative to the pre-period mean and are of a comparable relative magnitude to the effects on reporting to work. The effect on arrests, in contrast, is much less precise: we can rule out changes relative to the pre-period mean greater than .6, though the confidence intervals range between modest decreases and modest increases in the arrest rate. The differences in precision may be due to the relative rarity of arrests to calls for service, which most patrol officers will respond to at least once on a regular work day.

In Columns 1 and 2, Table 4 reports regression results for an indicator for any responses to 911 calls and the total volume of responses,<sup>7</sup> respectively. In both cases, we find relatively precise null results, though the estimate on counts in Column 2 is noisier. In Column 4, we again have report an insignificant result, though the estimate is much less precise in comparison.

In ??, we report our estimates for two additional enforcement outcomes, investigatory stops, similar to the “stop-and-frisk” policy in New York City, and use of force. Both outcomes, which are rare in the data, are imprecisely estimated. Although the point estimates imply a small or null result, the confidence intervals span nearly 50% of the pre-period mean in either direction in the case of stops. The estimates on the

---

<sup>7</sup>NOTE ON TOP-CODING

probability of using force are significantly noisier. Regressions for these two outcomes are reported in Columns 3 and 6 of Table 4, respectively. Even pooling the post-period observations, the estimated effect on use of force in Column 6 imply a wide range of possible effects, from a nearly .8 decline to a .22 increase.

Although we are unable to say much about the effects of RDO cancellation on rare enforcement outcomes, we can further investigate officer call response times. Despite the large volume of data on calls for service, most observations are missing timestamp information on at least one part of the response process. We therefore test the effects of RDO cancellations on four different measures. In Figure 12b, we report event study plots for the effects on the number of seconds passing between the time that a 911 call dispatcher puts out a radio call for officers to respond until the time that an officer reports that they have arrived on-scene. With the exception of event time 1, we can rule out modest changes in response times as a result of treatment. We do, however, see a statistically and economically significant increase in response times on the day immediately following an RDO cancellation, ranging from about one to three minute increases. In Figure 13b, we instead consider the time between when an officer is first dispatched to a call and when they arrive on-scene. Though this outcome is the best measure of a 911 response time, there are many fewer observations with both time stamps relative to the time from when a 911 dispatcher receives a call to when the officer arrives on-scene. We find qualitatively similar results as in Figure 12b, with possibly null effects after event time 1 but a point estimate of 100 seconds on the day after treatment. The confidence intervals are much larger in comparison, however.

In Table 5, we report regression estimates for these two response time measures in addition to two more, time from when the call is dispatched to when the case is cleared and time from when an officer acknowledges a call to when the case is cleared. The estimates vary widely across the four columns, in part because they measure different objects with different sample sizes. Nevertheless, in relative magnitudes, we find mostly precise null results. While we can, depending on the outcome, rule out modest and even small changes in the pre-period mean over the full five days following treatment, patterns in the event study plots suggest that there may be a real increase in response times on the day immediately following an RDO cancellation. The interpretation of this single significant coefficient is not straightforward, however. Although treatment and control officers do not report differential absence rates on the day after treatment, this day is technically a regularly scheduled day off. Therefore, officers reporting to work on event time 1 will necessarily be working voluntary overtime hours, up to measurement error in officer assignments. Not only will these officers be unrepresentative of the typical patrol officers, but many are likely to be on their sixth consecutive day of work without rest. It is therefore unclear to what extent this estimate can be explained by selection, treatment heterogeneity, an interaction with recent overtime or number of consecutive days worked.

Given the rarity of many of our outcomes, we also test whether we are able to detect effects on aggregations of different variables. In Column 1 of ??, we estimate

the effect of treatment on the probability that an officer either calls in sick or is absent due to an on-the-job injury. We find little evidence of an effect with levels of precision comparable to the results in Columns 1 and 2 in Table 3. In Columns 2 and 3, we then consider aggregations across all enforcement actions in our data, including 911 call responses, arrests, stops, and use of force. In Column 2, we find a relatively precise null effect on an indicator equal to one of an officer reports any of the aforementioned actions. In Column 3, we then consider the total sum of all enforcement actions reported by an officer each day: total 911 response, arrest, stop, and use of force counts. We again find an insignificant result and can rule out modest decreases in the count of enforcement activities.

## 5.5 Treatment Heterogeneity

In this section, we next test for heterogeneity in the effects of mandatory overtime. It is possible, for instance, that heterogeneity in officer productivity may mask underlying treatment effects. If some officers are less productive, then exhaustion from an RDO cancellation may not impact their enforcement outcomes, whereas we may see important effects on more productive officers. In the following tables, we test this hypothesis by using one measure of officer productivity, namely the average daily number of calls for service an officer responded to in 2021. To assess the presence of heterogeneity with respect to this measure of productivity, we estimate the median of average call volumes and compare officers above and below the median.

Table B.20 reports regressions of work outcomes on triple-difference interactions between indicators for treatment, post-treatment, and being above the median response time in 2021. For all outcomes, the estimated coefficients on the double- and triple differences are statistically insignificant. For the main difference-in-difference interactions, the 95% confidence intervals are comparable to the results in Table 3. For the triple interactions, we can rule out small- to medium-sized effects relative to the pre-period mean, depending on the outcome. The evidence suggests there is not an economically significant difference in the estimated effects on working and absences between more and less productive officers.

In Table B.21, we next turn to enforcement outcomes. The estimated coefficients on the double interactions are comparable to the estimates in ??, with the confidence intervals on arrests and use of force suggesting that moderate effects cannot be ruled out with 95% confidence. The triple interactions follow a similar pattern: we can preclude economically important effects on call for service responses and investigatory stops but not on arrests and the use of force. In Table B.22, we then test for heterogeneity in different measures of the time an officer takes to respond to a call for service. We find little evidence of moderate heterogeneous effects on response times with respect to officer productivity, with some estimates being precise enough to rule out smaller effects with 95% confidence.

It is also possible that the effects of treatment are changing over time. For example, if officer outcomes and productivity are reference-dependent, initial RDO cancellations may have significant negative effects. As officers adapt, later cancellations may have few effects. Alternatively, officers may be able to weather the first few cancellations. As they get repeatedly exposed to treatment, however, work exhaustion can build overtime and induce officer burnout. To test these hypotheses, we find that the median number of times officers in our treatment group are assigned to an RDO cancellation is thirteen. As in Table B.20, we regress our outcomes on a triple-difference interacting the difference-in-differences term with an indicator for whether an officer has been treated thirteen or more times thus far.

In Table B.23, we again consider a range of indicators for officer work attendance and absence. As before, we find little evidence of economically and statistically significant treatment heterogeneity. A possible exception is the estimated coefficient on the double interaction for on-the-job injuries. Officers exposed to treatment fewer times than the median are about 2% more likely to be absent due to a work injury. Although the estimate is significant at the 10% level, we can rule out effects larger than 3% of the pre-period mean with 95% confidence, a relatively small effect. Moreover, this single marginally significant estimate is unlikely to affect a joint significance test and finding it may be the result of multiple hypothesis testing.

Next, Table B.24 presents estimates for treatment heterogeneity with respect to the number of times treated on our main enforcement outcomes. The findings are qualitatively similar to the results in Table B.21, with the estimates on arrests and use of force too noisy to rule out economically significant effects. For response times, however, Table B.25 there appears to be some evidence of treatment heterogeneity. Specifically, in Column 1, our outcome is the time it takes for an officer to arrive on-scene from the time that a call dispatcher receives a call for service. Here, we find statistically significant evidence that that RDO cancellations may delay response times up to about a minute early on in the policy. Though small relative to a mean response time of about 13 and a half minutes, this effect may be relevant for public safety in urgent situations. Treatment exposure later on appears to have a null effect on response times: the coefficient on the triple interaction, which is estimated with a similar level of precision as the double interaction, is negative and approximately cancels out the main effect. The range of effect sizes in the 95% confidence intervals also appear to cancel each other out. We find similar patterns in Columns 2 and 4, though only the double interaction on the latter is estimated precisely. In Column 3, we instead find positive effects for both interactions, though the effects are statistically insignificant. Despite being mostly insignificant, the estimates in Table B.25 are precise enough overall to rule out moderate changes in response times.

Although the median number of times assigned to treatment may be a reasonable dimension of heterogeneity, it is possible that officers adjust to the policy sooner than their thirteenth exposure. In results not presented here but available upon request, we also test for heterogeneity with respect to the second, fifth, and tenth time that an

officer is assigned to treatment. Each case produces similar findings, with little evidence overall of economically or statistically relevant treatment heterogeneity.

## 6 Conclusion

In this paper, we evaluate the effects of the Chicago Police Department’s mandatory overtime policy on officer well-being and productivity. Leveraging cross-sectional and temporal variation in day-off cancellations through a stacked difference-in-differences design, we find that overall, the policy did not significantly affect officer-level outcomes.

We first show that the policy was successfully implemented: officers respond to treatment, though with imperfect compliance. Moreover, we do not find evidence that officers evade treatment or undo the effect of the policy by taking additional days off. The precision of our estimates allows us to rule out small differences in absence rates between treatment and control officers. We also find no evidence of negative health effects on officers as a result of the policy and can rule out modest changes in illness-related absences. On enforcement outcomes, point estimates imply small and potentially zero effects, but rare enforcement outcomes estimated imprecisely. We can rule out moderate changes in the volume of responses to calls for service, the most common enforcement outcome in the data. Although we can also rule out small changes in call response times over the full post-treatment period, we do find meaningful and significant increases in response times on the day immediately following treatment. Overall, the risk of the policy appears to be low, especially relative to a setting in which voluntary overtime is concentrated among a small group of officers while the CPD cannot directly manage the distribution of overtime hours.

We highlight two caveats to our analysis. First, we are only able to identify the short-run effects of mandatory overtime. Given limits in the data, we can only extend our analysis out to two full work cycles after treatment. Because most officers under this policy were assigned to mandatory overtime no more than once a month, it is not an unreasonable conjecture that most of the effects of a day-off cancellation will play out within the following two weeks. In contrast, we are limited in what we can say about the cumulative effects of this policy over time. For example, if the CPD were to cancel a day off for all officers every a month moving forward, repeated instances of involuntary overtime over an extended period of time may induce burnout, affecting officer well-being, productivity, retainment, and possibly public safety.

The second caveat is that we can only assess the effect of a single day-off cancellation. Consider a policy that cancels two days off in a row for the officers in our sample. Given the “four days working, two days off” pattern in work schedules, such officers would be assigned ten consecutive days of work. The effects of such a policy are unlikely to be as modest as what we find for a single day off cancellation. Our only consistent and unambiguous significant result, an increase in call response times on the day after treatment, highlights this point: officers normally have the day after treatment off, so officers responding to calls for service on the following day are likely to be on their sixth

consecutive day of work. Despite these caveats, our findings suggest that a moderate mandatory overtime policy may be a useful short-run tool for police departments and other organizations facing significant labor constraints.

## References

- Ba, B. A., D. Knox, J. Mummolo, and R. Rivera (2021). The role of officer race and gender in police-civilian interactions in chicago. *371*(6530), 696–702. Publisher: American Association for the Advancement of Science.
- Blackburn, B., T. Chan, E. Cherot, R. B. Freeman, X. Hu, E. Matt, and C. A. Rhodes (2023, January). Beyond burnout: From measuring to forecasting. NBER Working Paper Series 30895, National Bureau of Economic Research. Working Paper.
- Cho, S., F. Gonçalves, and E. Weisburst (2023, June). The impact of fear on police behavior and public safety. NBER Working Paper Series 31392, National Bureau of Economic Research. Working Paper.
- City of Chicago Office of Inspector General (2022, August). Consecutive days worked by cpd members, april-may 2022.
- DeAngelo, G., M. Toger, and S. Weisburd (2023). Police response time and injury outcomes. *The Economic Journal* *133*(August), 2147–2177.
- Doan, S., E. D. Steiner, and R. Pandey (2024, June). Teacher well-being and intentions to leave in 2024: Findings from the 2024 state of the american teacher survey. Research Report RRA1108-12, RAND Corporation.
- Ehrenberg, R. G. (1970, June). Absenteeism and the overtime decision. *The American Economic Review* *60*(3), 352–357.
- Ferrazares, T. (2025). Shift structure and cognitive depletion: Evidence from police officers. Working Paper. UC Santa Barbara. <https://ferrazares.github.io/assets/FERRAZARESjmp.pdf>.
- Gudgeon, M., A. Jordan, and T. Kim (2024). Do teams perform differently under black and hispanic leaders? evidence from the chicago police department. Working Paper. Tufts University. <https://drive.google.com/file/d/1jn9a5QxE6Up48JCAKqummGF7pRbeUu6F/view?pli=1>.
- Holz, J. E., R. G. Rivera, and B. A. Ba (2023). Peer effects in police use of force. *15*(2), 256–291.
- International Association of Chiefs of Police (2020). The state of recruitment: A crisis for law enforcement. Technical report, International Association of Chiefs of Police, Alexandria, VA.
- James, L. (2018). The stability of implicit racial bias in police officers. *21*(1), 30–52.
- Linos, E., K. Ruffini, and S. Wilcoxon. Reducing burnout and resignations among frontline workers: A field experiment.
- Maryland Office of Legislative Audits (2024, January). Baltimore police department: A review of overtime policies, procedures, and activity. Performance audit report, Department of Legislative Services, Maryland General Assembly.
- National Conference of State Legislatures (2021). Law enforcement legislation: Significant trends 2021. Technical report, National Conference of State Legislatures. Analysis of state legislative responses to law enforcement policy following 2020 events.
- Nekoei, A., J. Sigurdsson, and D. Wehr (2024, May). The economic burden of burnout. Working Paper 11128, CESifo. [https://www.cesifo.org/DocDL/cesifo1\\_wp11128.pdf](https://www.cesifo.org/DocDL/cesifo1_wp11128.pdf).

- Pandemic Response Accountability Committee (2023, September). Review of personnel shortages in federal health care programs during the covid-19 pandemic. Technical report, Pandemic Response Accountability Committee. Collaborative report by DOD OIG, DOJ OIG, VA OIG, and HHS OIG.
- Police Executive Research Forum (2023). New perf survey shows police agencies are losing officers faster than they can hire new ones. Accessed: 2024-08-21.
- Sachiko, K. and Y. Isamu (2016, March). Does mental health matter for firm performance? evidence from longitudinal japanese firm data. Working Paper 16-E-016, The Research Institute of Economy, Trade and Industry. <https://www.rieti.go.jp/jp/publications/dp/16e016.pdf>.
- San José Office of the City Auditor (2016, September). Police overtime: The san josé police department relies on overtime to patrol the city due to unprecedented vacancies. Technical Report Report 16-08, City of San José. Report to the City Council.
- Sisk, C. (2023, October). New efforts to boost officer morale in cpd gives time back to officers. CBS Chicago, accessed September 11, 2024.
- Stoughton, S. W. (2016). Moonlighting: The private employment of off-duty officers.
- U.S. Census Bureau (2024). U.s. population clock. Accessed: 2024-08-21.
- Whitaker, S. (2023, August). The costs of police overtime in charleston. Issue brief, West Virginia Center on Budget and Policy.
- Wing, C., S. M. Freedman, and A. Hollingsworth (2023). Stacked difference-in-differences. *National Bureau of Economic Research Working Paper Series* (32054).
- Yaniv, G. (1995). Burnout, absenteeism, and the overtime decision. *Journal of Economic Psychology* 16(2), 297–309.



## 7 Tables

### 7.1 Descriptive Statistics

**Table 1:** Summary Statistics for Analytical Sample

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Distinct Officers	7,566	-	-	-	-
Reports to Work	2,438,867	0.539	0.499	0	1
Any Involuntary Overtime Hours	2,438,867	0.0231	0.150	0	1
Total Overtime Hours	2,438,867	0.901	2.516	0	34.50
Total Involuntary Overtime Hours	2,438,867	0.199	1.303	0	25
Total Voluntary Overtime Hours	2,438,867	0.703	2.197	0	34.50
Actual DOG Changed	2,438,867	0.00286	0.0534	0	1
Actual Area Changed	2,438,867	0.00211	0.0459	0	1
Injured on the Job	2,432,108	0.0357	0.186	0	1
Calls in Sick	2,432,108	0.0580	0.234	0	1
Paid Time Off	2,432,108	0.0807	0.272	0	1
Total Calls for Service	2,438,867	1.887	3.510	0	21
Total Investigatory Stops	2,438,867	0.0439	0.312	0	12
Total Arrests	2,438,867	0.0709	0.319	0	10
Any Use of Force	2,438,867	0.00110	0.0331	0	1

**Table 2:** Summary Statistics for Stacked Sample

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Distinct Officers	5,606	-	-	-	-
Distinct Sub-Experiments	215	-	-	-	-
Officers per Sub-Experiment	215	576.0	151.6	233	876
Treated	1,261,443	0.384	0.486	0	1
Reports to Work	1,261,443	0.539	0.498	0	1
Any Involuntary Overtime Hours	1,261,443	0.0200	0.140	0	1
Total Overtime Hours	1,261,443	0.746	2.258	0	26
Total Involuntary Overtime Hours	1,261,443	0.172	1.216	0	25
Total Voluntary Overtime Hours	1,261,443	0.574	1.937	0	26
Actual DOG Changed	1,261,443	0.0227	0.149	0	1
Actual Area Changed	1,261,443	0.0110	0.104	0	1
Injured on the Job	1,257,346	0.0450	0.207	0	1
Calls in Sick	1,257,346	0.0648	0.246	0	1
Paid Time Off	1,257,346	0.0940	0.292	0	1
Total Calls for Service	1,261,443	2.362	3.853	0	21
Total Investigatory Stops	1,261,443	0.0573	0.356	0	12
Total Arrests	1,261,443	0.0291	0.209	0	8
Any Use of Force	1,261,443	0.00164	0.0405	0	1

## 7.2 Baseline Regressions

**Table 3:** Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0075 (0.0031) {-2.4066}	-0.0002 (0.0016) {-0.1318}	0.0004 (0.0010) {0.4476}	0.0001 (0.0004) {0.3773}	0.0017 (0.0018) {0.9580}
95% CI	[-0.0136, -0.0014]	[-0.0034, 0.0030]	[-0.0014, 0.0023]	[-0.0006, 0.0008]	[-0.0018, 0.0052]
CI Rel. to Pre-Per. Mean	[-2.37%, -0.24%]	[-2.92%, 2.55%]	[-2.30%, 3.66%]	[-1.28%, 1.89%]	[-1.81%, 5.27%]
CI Rel. to Pre-Per. SD	[-2.74%, -0.28%]	[-1.06%, 0.92%]	[-0.59%, 0.95%]	[-0.28%, 0.41%]	[-0.60%, 1.75%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1685760	1685760	1682220	1682220	1682220

**Table 4:** Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0049 (0.0026) {-1.8959}	-0.0049 (0.0026) {-1.8959}	-0.0012 (0.0009) {-1.4610}	-0.0024 (0.0017) {-1.3635}	-0.0001 (0.0002) {-0.4410}
95% CI	[-0.0100, 0.0002]	[-0.0100, 0.0002]	[-0.0029, 0.0004]	[-0.0058, 0.0010]	[-0.0005, 0.0003]
CI Rel. to Pre-Per. Mean	[-2.26%, 0.04%]	[-2.26%, 0.04%]	[-7.29%, 1.06%]	[-7.17%, 1.29%]	[-25.58%, 16.18%]
CI Rel. to Pre-Per. SD	[-2.02%, 0.03%]	[-2.02%, 0.03%]	[-1.49%, 0.22%]	[-2.12%, 0.38%]	[-1.09%, 0.69%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1685760	1685760	1685760	1685760	1685760

**Table 5:** Calls for Service Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-4.8709 (7.0400) {-0.6919}	-10.9263 (10.0168) {-1.0908}	14.0002 (79.5709) {0.1759}	189.0513 (121.7143) {1.5532}
95% CI	[-18.6694, 8.9275]	[-30.5592, 8.7065]	[-141.9588, 169.9593]	[-49.5088, 427.6114]
CI Rel. to Pre-Per. Mean	[-3.59%, 1.72%]	[-5.60%, 1.60%]	[-2.49%, 2.98%]	[-0.71%, 6.10%]
CI Rel. to Pre-Per. SD	[-2.39%, 1.14%]	[-3.93%, 1.12%]	[-2.02%, 2.42%]	[-0.52%, 4.50%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	688516	342388	744025	508758

Table 6: Arrest Type

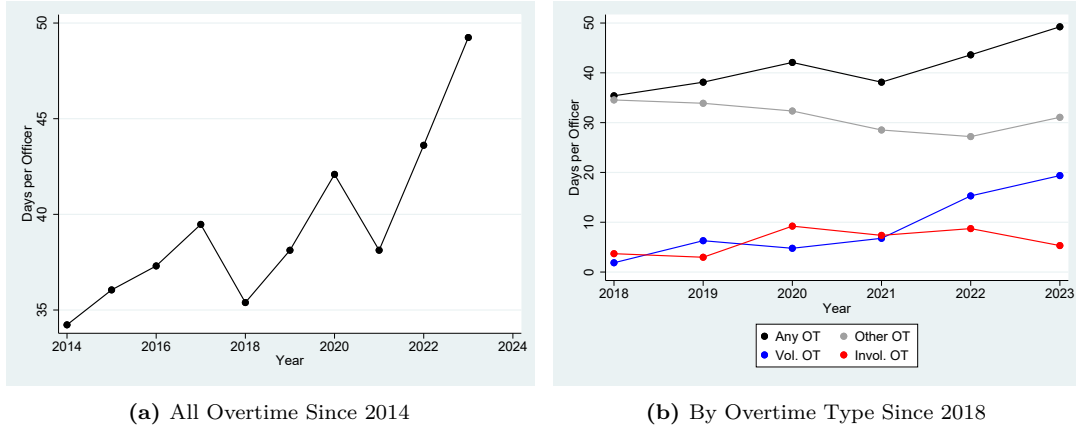
Dependent Variable:	Any Index Arrests	Any Traffic Arrests	Other Arrests	Any Arrests with No Charge
Treatment $\times$ Post	-0.0003 (0.0009) {-0.3171}	-0.0011 (0.0011) {-0.9926}	-0.0017 (0.0018) {-0.9422}	0.0001 (0.0004) {0.2550}
95% CI	[-0.0020, 0.0014]	[-0.0032, 0.0010]	[-0.0051, 0.0018]	[-0.0008, 0.0010]
CI Rel. to Pre-Per. Mean	[-12.07%, 8.71%]	[-13.35%, 4.37%]	[-7.29%, 2.56%]	[-24.05%, 31.25%]
CI Rel. to Pre-Per. SD	[-1.55%, 1.12%]	[-2.09%, 0.68%]	[-2.00%, 0.70%]	[-1.35%, 1.75%]
Pre-Period Mean	0.0163	0.0238	0.0699	0.0031
Pre-Period SD	0.1267	0.1525	0.2550	0.0559
Clusters	30	30	30	30
Observations	1685760	1685760	1685760	1685760

**Table 7:** Combined Outcomes

Dependent Variable:	Sick or Injured	Any Enforcement Action	Count of Enforcement Actions
Treatment $\times$ Post	0.0006 (0.0010) {0.5713}	-0.0057 (0.0028) {-2.0151}	-0.0111 (0.0049) {-2.2430}
95% CI	[-0.0014, 0.0025]	[-0.0112, -0.0002]	[-0.0207, -0.0014]
CI Rel. to Pre-Per. Mean	[-1.30%, 2.36%]	[-2.42%, -0.03%]	[-3.39%, -0.23%]
CI Rel. to Pre-Per. SD	[-0.45%, 0.82%]	[-2.25%, -0.03%]	[-2.44%, -0.16%]
Pre-Period Mean	0.1069	0.4627	0.6103
Pre-Period SD	0.3090	0.4986	0.8491
Clusters	30	30	30
Observations	1682220	1685760	1685760

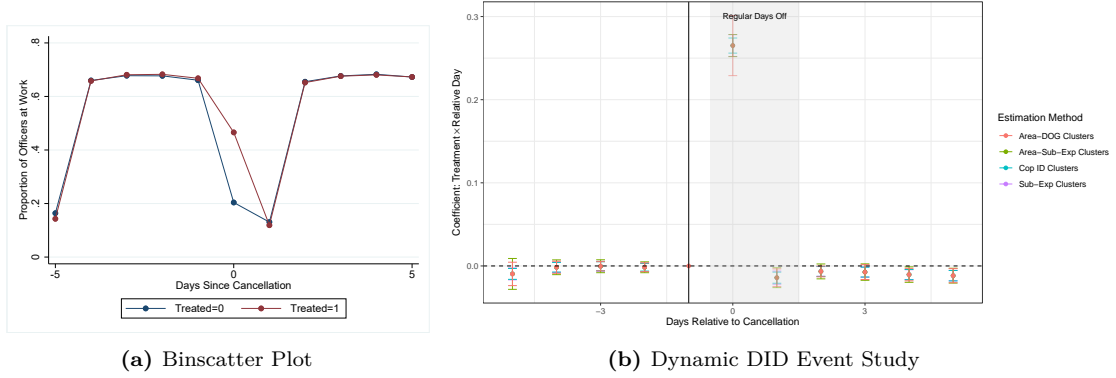
## 8 Figures

### 8.1 Overtime Descriptive Statistics

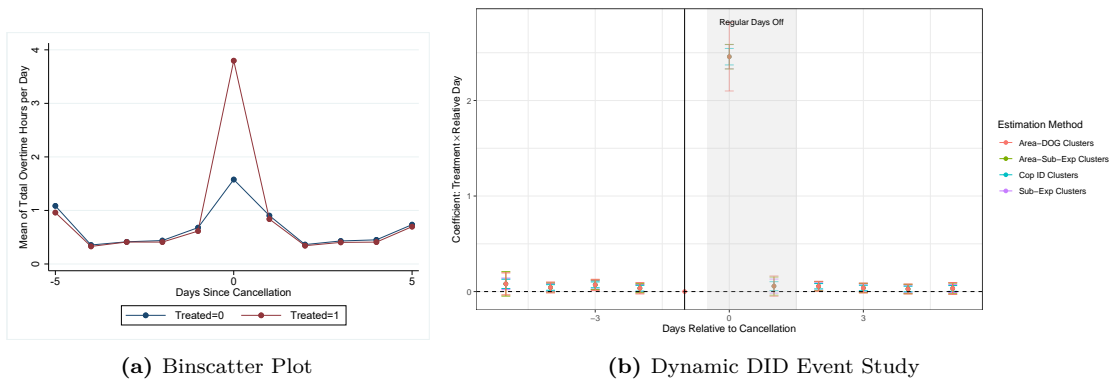


**Figure 1:** Average Days Working Overtime per Officer

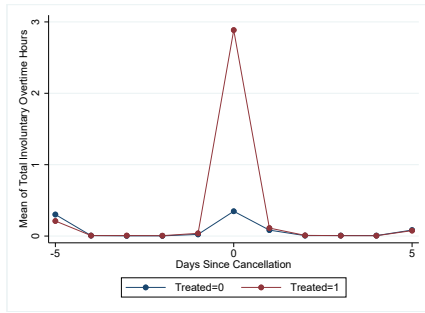
### 8.2 Effects on Attendance



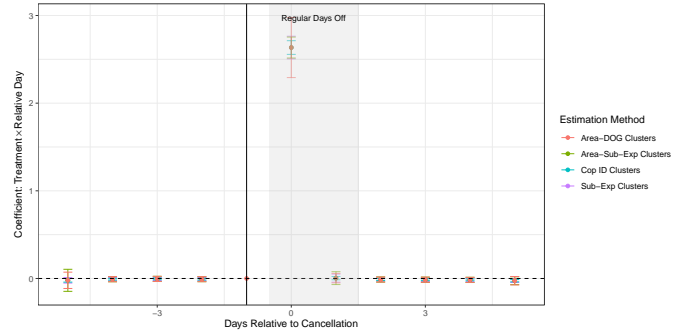
**Figure 2:** Officer Reports to Work



**Figure 3:** Total Overtime Hours

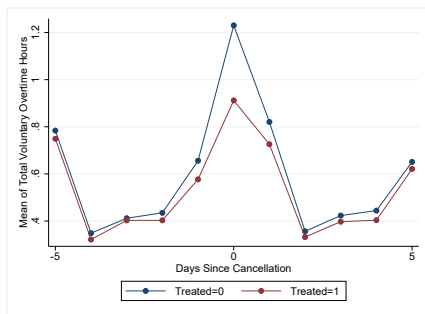


(a) Binscatter Plot

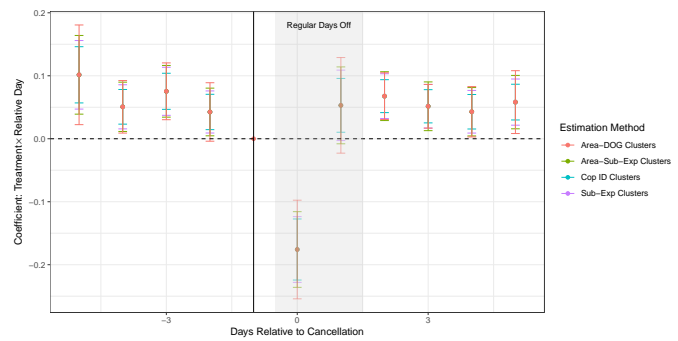


(b) Dynamic DID Event Study

**Figure 4: Involuntary Overtime Hours**



(a) Binscatter Plot



(b) Dynamic DID Event Study

**Figure 5: Voluntary Overtime Hours**



### 8.3 Effects on Officer Wellness

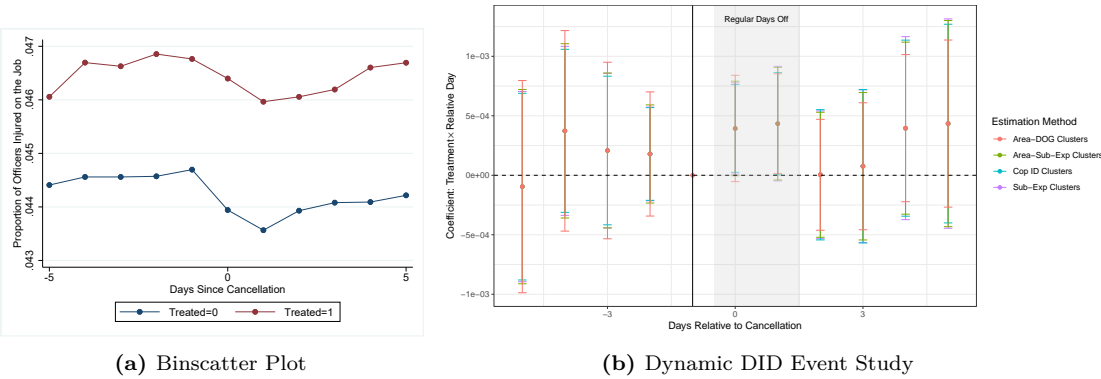


Figure 6: Injured on the Job

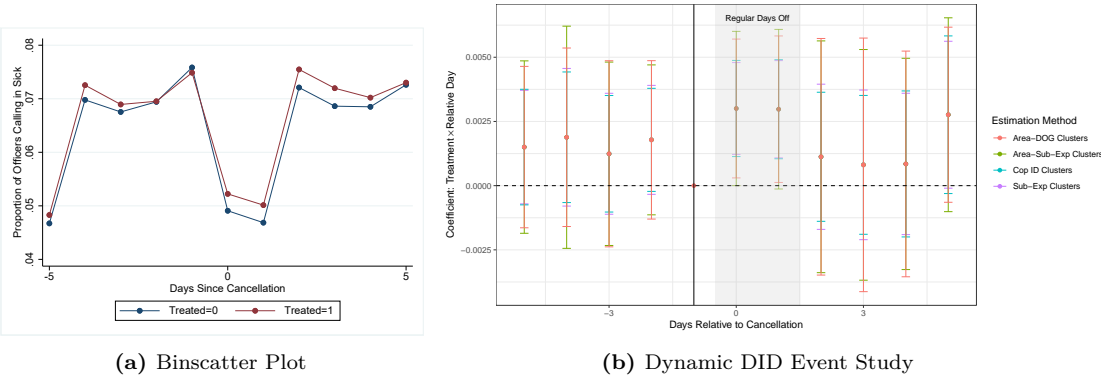
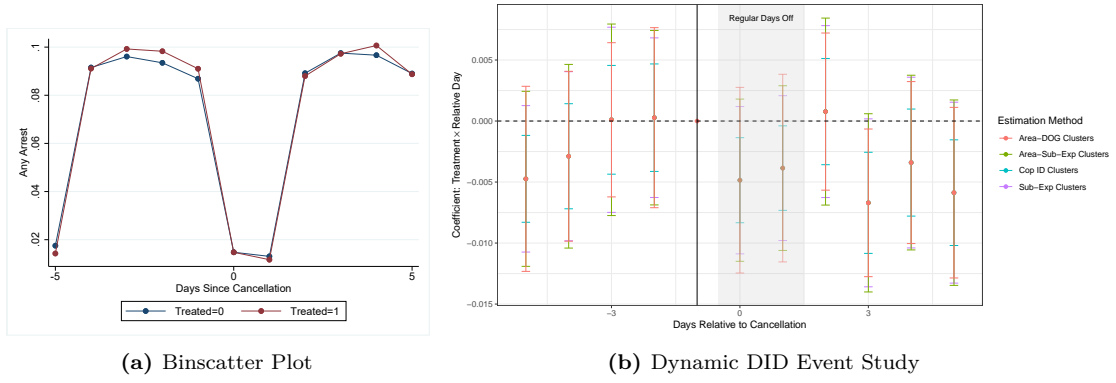
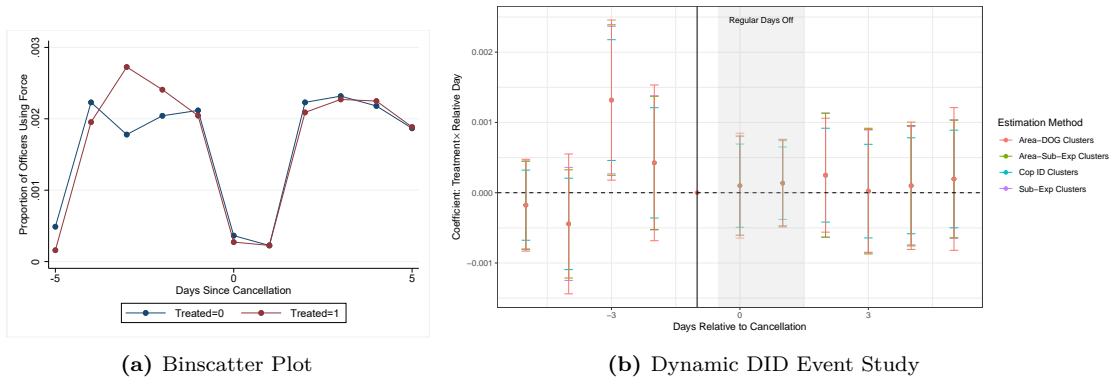


Figure 7: Calls in Sick

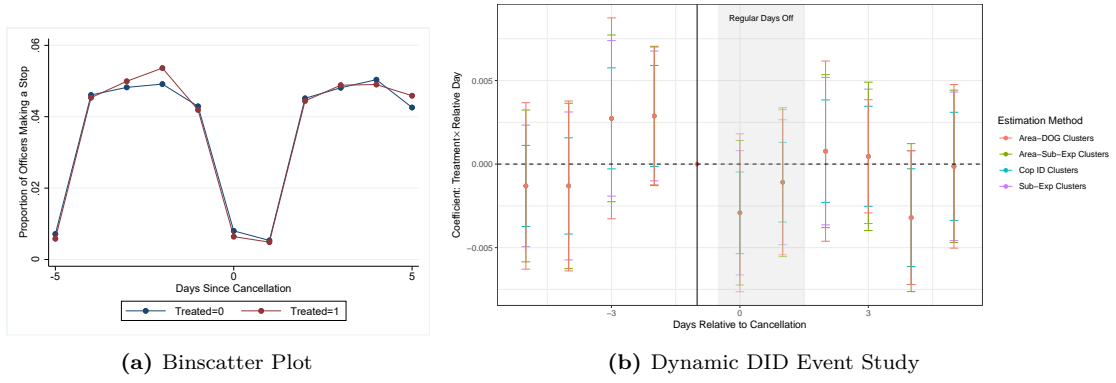
## 8.4 Effects on Law Enforcement Activity



**Figure 8: Any Arrests**

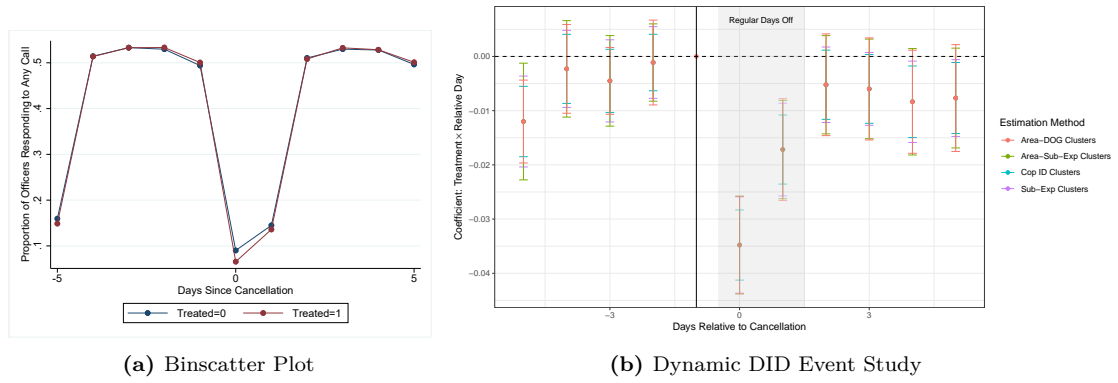


**Figure 9: Use of Force**

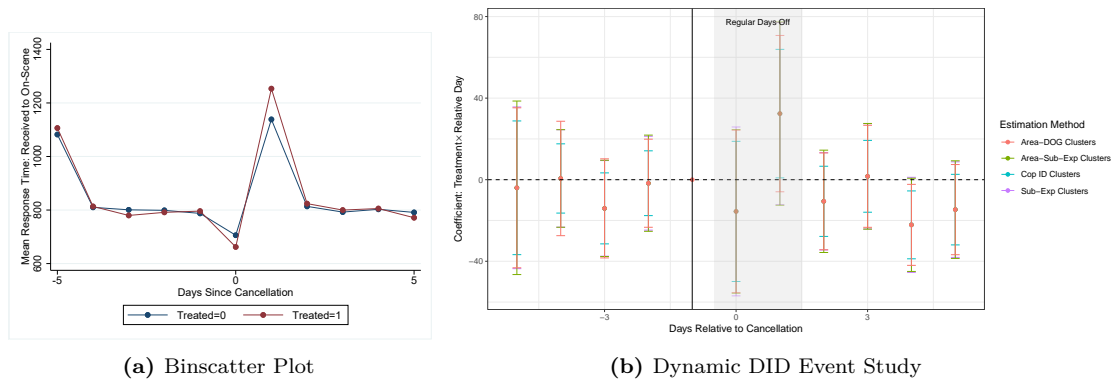


**Figure 10: Any Investigatory Stop**

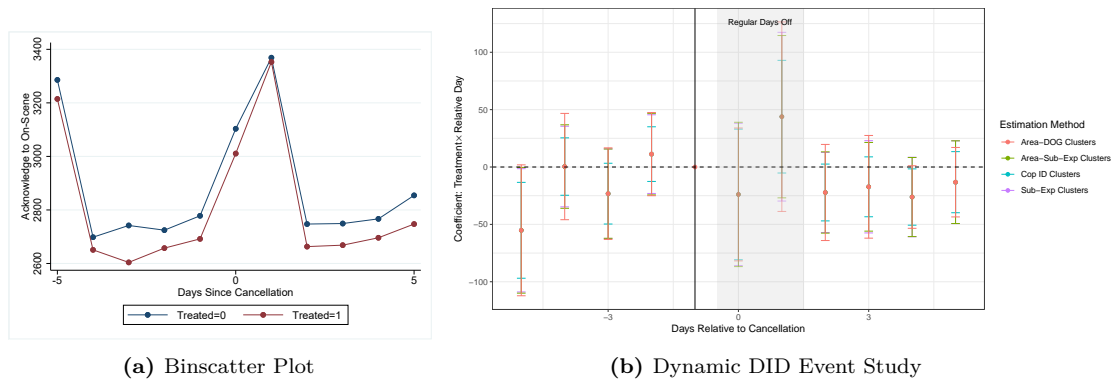
## 8.5 Effects on Officer Productivity



**Figure 11:** Responds to Any Call for Service



**Figure 12:** Time from Call Received by Dispatcher to On-Scene in Seconds



**Figure 13:** Time from Officer Dispatched to On-Scene in Seconds

## 8.6 Treatment Heterogeneity by Officer Characteristics

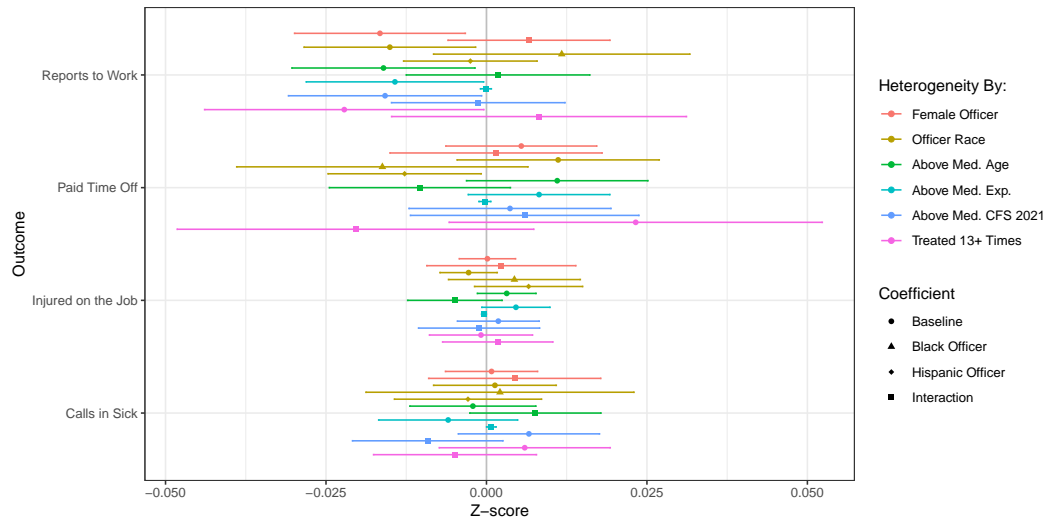


Figure 14: Attendance and Wellness

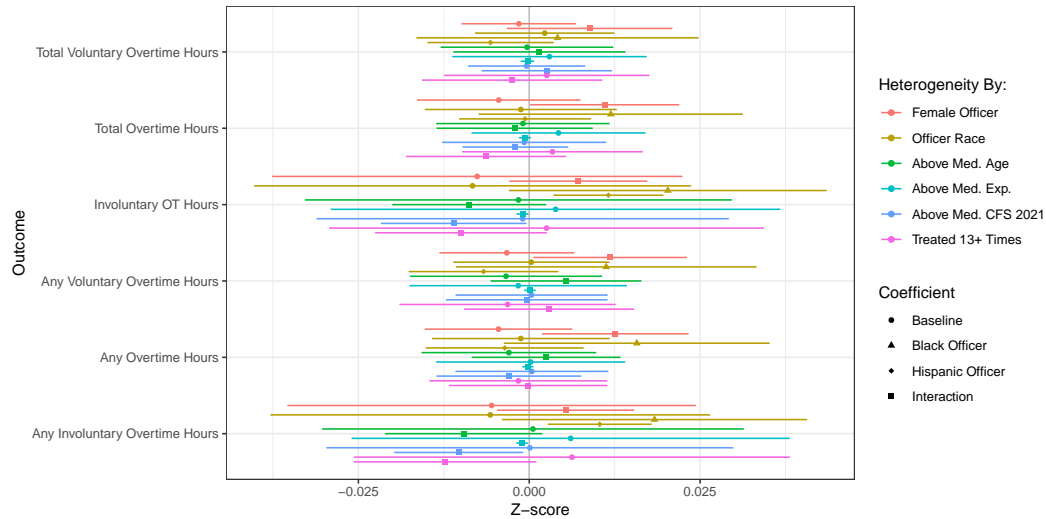
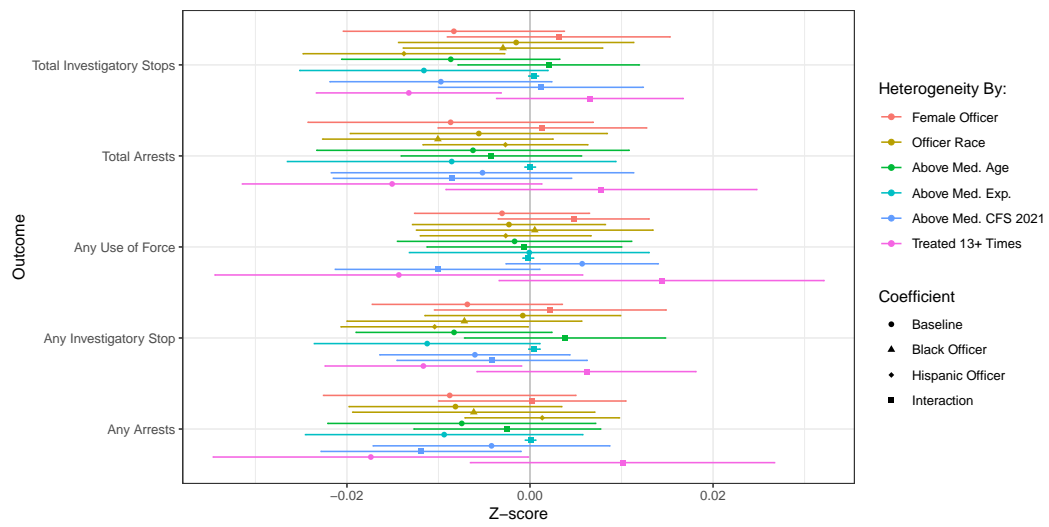
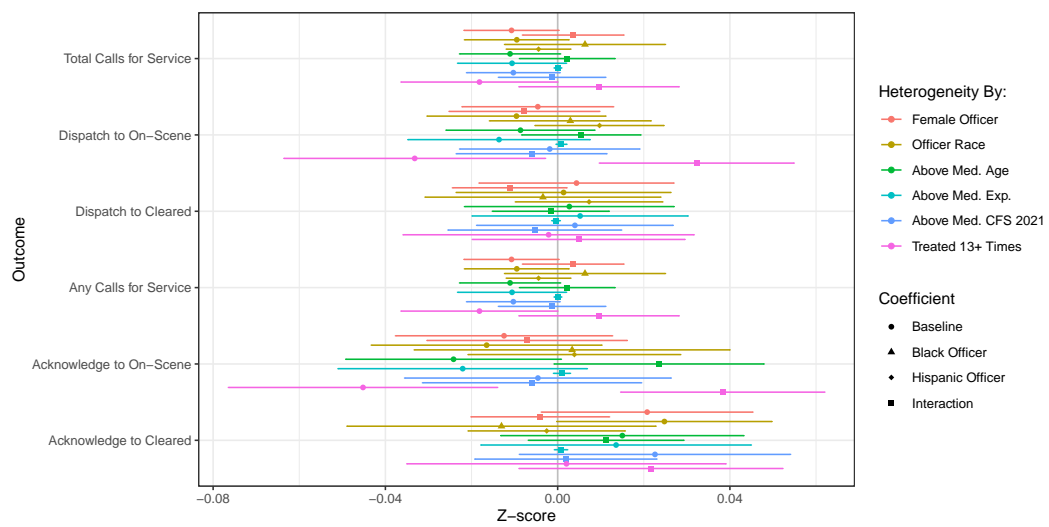


Figure 15: Overtime



**Figure 16: Enforcement**



**Figure 17: Calls for Service**

## A Data

We analyze data from CPD officers in DOGs 61-66, which includes approximately 66% of officer-date combinations we observe in a linked A&A and officer roster data set. We restrict our sample to this subset of DOGs because this

We also restrict our sample to officers assigned areas numbered 1-5. While The CPD does assign some officers to “areas” with number labels greater than 5, these assignments are mainly administrative, are not associated with a specific geographic area, and so the job tasks fall outside our productivity measures of interest.

We identify RDO cancellations using cancellation matrices obtained from the CPD. Figure X shows an example of an RDO cancellation matrix we obtained from the CPD through a FOIA request. Figure X shows the sequential, cyclic structure to RDO cancellations, with occasional exceptions. For example, on May 31, 2022, the cancellation cycle begins with officers in Area 1 and DOG 61<sup>8</sup>. Comparing this figure to the operations calendar for 2022 in Figure X, it can be seen that May 31st is the first of two regularly scheduled days-off for DOG 61. The following day, June 1st, 2022, officers in Area 1 in DOG 62 were assigned an RDO cancellation and so on until by June 5th, 2022, all officers in DOGs 61 to 66 in Area 1 had been assigned treatment. Next, on June 7th, officers in all DOGs 61 to 66 in Areas 3 and 4 were sequentially assigned treatment, starting with DOG 62.

---

<sup>8</sup>DOG 75 also experiences an RDO cancellation on this date, but our analysis is restricted to officers in DOGs 61 to 66. See 3 for details on this sample restriction.

## B Tables

# B.1 Heterogeneity by Officer Gender

**Table B.8:** Heterogeneity by Gender: Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0082 (0.0034) {-2.4447}	-0.0011 (0.0016) {-0.6581}	0.0002 (0.0009) {0.2158}	0.0000 (0.0005) {0.0657}	0.0016 (0.0018) {0.9042}
95% CI	[-0.0148, -0.0016]	[-0.0042, 0.0021]	[-0.0015, 0.0019]	[-0.0009, 0.0009]	[-0.0019, 0.0051]
CI Rel. to Pre-Per. Mean	[-2.59%, -0.28%]	[-3.61%, 1.79%]	[-2.47%, 3.08%]	[-1.97%, 2.11%]	[-1.91%, 5.17%]
CI Rel. to Pre-Per. SD	[-2.99%, -0.33%]	[-1.31%, 0.65%]	[-0.64%, 0.80%]	[-0.42%, 0.45%]	[-0.63%, 1.72%]
Treatment $\times$ Post $\times$ Female	0.0033 (0.0032) {1.0289}	0.0038 (0.0018) {2.0805}	0.0011 (0.0017) {0.6439}	0.0005 (0.0012) {0.3911}	0.0004 (0.0025) {0.1771}
95% CI	[-0.0030, 0.0095]	[0.0002, 0.0074]	[-0.0022, 0.0043]	[-0.0019, 0.0029]	[-0.0045, 0.0054]
CI Rel. to Pre-Per. Mean	[-0.52%, 1.67%]	[0.19%, 6.35%]	[-3.48%, 6.89%]	[-4.32%, 6.47%]	[-4.52%, 5.42%]
CI Rel. to Pre-Per. SD	[-0.60%, 1.93%]	[0.07%, 2.30%]	[-0.90%, 1.78%]	[-0.93%, 1.39%]	[-1.50%, 1.80%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1685760	1685760	1682220	1682220	1682220



**Table B.9:** Heterogeneity by Gender: Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0053 (0.0028) {-1.9040}	-0.0053 (0.0028) {-1.9040}	-0.0013 (0.0010) {-1.2932}	-0.0024 (0.0019) {-1.2473}	-0.0001 (0.0002) {-0.6271}
95% CI	[-0.0108, 0.0002]	[-0.0108, 0.0002]	[-0.0034, 0.0007]	[-0.0061, 0.0014]	[-0.0005, 0.0003]
CI Rel. to Pre-Per. Mean	[-2.44%, 0.04%]	[-2.44%, 0.04%]	[-8.43%, 1.73%]	[-7.61%, 1.69%]	[-29.63%, 15.26%]
CI Rel. to Pre-Per. SD	[-2.17%, 0.03%]	[-2.17%, 0.03%]	[-1.72%, 0.35%]	[-2.26%, 0.50%]	[-1.26%, 0.65%]
Treatment $\times$ Post $\times$ Female	0.0018 (0.0030) {0.6025}	0.0018 (0.0030) {0.6025}	0.0004 (0.0013) {0.3445}	0.0001 (0.0014) {0.0477}	0.0002 (0.0002) {1.1359}
95% CI	[-0.0040, 0.0076]	[-0.0040, 0.0076]	[-0.0020, 0.0029]	[-0.0027, 0.0029]	[-0.0001, 0.0006]
CI Rel. to Pre-Per. Mean	[-0.91%, 1.72%]	[-0.91%, 1.72%]	[-5.11%, 7.28%]	[-3.37%, 3.54%]	[-8.15%, 30.60%]
CI Rel. to Pre-Per. SD	[-0.81%, 1.54%]	[-0.81%, 1.54%]	[-1.04%, 1.49%]	[-1.00%, 1.05%]	[-0.35%, 1.30%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1685760	1685760	1685760	1685760	1685760

**Table B.10:** Heterogeneity by Gender: Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-3.5958 (7.0002) {-0.5137}	-9.6787 (9.9878) {-0.9691}	30.7333 (81.0410) {0.3792}	197.4495 (118.9588) {1.6598}
95% CI	[-17.3163, 10.1246]	[-29.2548, 9.8974]	[-128.1072, 189.5737]	[-35.7098, 430.6087]
CI Rel. to Pre-Per. Mean	[-3.33%, 1.95%]	[-5.37%, 1.82%]	[-2.25%, 3.32%]	[-0.51%, 6.14%]
CI Rel. to Pre-Per. SD	[-2.22%, 1.30%]	[-3.76%, 1.27%]	[-1.82%, 2.70%]	[-0.38%, 4.53%]
Treatment $\times$ Post $\times$ Female	-6.0158 (6.9577) {-0.8646}	-5.5089 (9.2171) {-0.5977}	-78.1240 (47.6062) {-1.6410}	-38.4216 (77.8787) {-0.4934}
95% CI	[-19.6528, 7.6213]	[-23.5745, 12.5567]	[-171.4322, 15.1843]	[-191.0638, 114.2207]
CI Rel. to Pre-Per. Mean	[-3.78%, 1.46%]	[-4.32%, 2.30%]	[-3.00%, 0.27%]	[-2.72%, 1.63%]
CI Rel. to Pre-Per. SD	[-2.52%, 0.98%]	[-3.03%, 1.62%]	[-2.44%, 0.22%]	[-2.01%, 1.20%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	688516	342388	744025	508758

**B.2 Heterogeneity by Officer Race**

**Table B.11:** Heterogeneity by Race: Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0074 (0.0034) {-2.2069}	0.0001 (0.0019) {0.0515}	0.0003 (0.0012) {0.2704}	-0.0006 (0.0005) {-1.2219}	0.0033 (0.0024) {1.3902}
95% CI	[-0.0141, -0.0008]	[-0.0035, 0.0037]	[-0.0020, 0.0026]	[-0.0015, 0.0003]	[-0.0014, 0.0081]
CI Rel. to Pre-Per. Mean	[-2.46%, -0.15%]	[-3.05%, 3.21%]	[-3.19%, 4.21%]	[-3.37%, 0.78%]	[-1.38%, 8.10%]
CI Rel. to Pre-Per. SD	[-2.84%, -0.17%]	[-1.10%, 1.16%]	[-0.82%, 1.09%]	[-0.72%, 0.17%]	[-0.46%, 2.69%]
Treatment $\times$ Post $\times$ Black	0.0058 (0.0050) {1.1510}	0.0036 (0.0036) {1.0079}	0.0005 (0.0026) {0.1975}	0.0009 (0.0011) {0.8311}	-0.0049 (0.0035) {-1.4007}
95% CI	[-0.0041, 0.0157]	[-0.0034, 0.0106]	[-0.0046, 0.0056]	[-0.0012, 0.0030]	[-0.0116, 0.0019]
CI Rel. to Pre-Per. Mean	[-0.71%, 2.74%]	[-2.94%, 9.16%]	[-7.26%, 8.89%]	[-2.75%, 6.79%]	[-11.71%, 1.95%]
CI Rel. to Pre-Per. SD	[-0.82%, 3.17%]	[-1.06%, 3.32%]	[-1.88%, 2.30%]	[-0.59%, 1.46%]	[-3.89%, 0.65%]
Treatment $\times$ Post $\times$ Hispanic	-0.0012 (0.0026) {-0.4723}	-0.0021 (0.0018) {-1.2068}	-0.0007 (0.0014) {-0.4923}	0.0014 (0.0009) {1.5278}	-0.0038 (0.0018) {-2.0941}
95% CI	[-0.0064, 0.0039]	[-0.0056, 0.0013]	[-0.0035, 0.0021]	[-0.0004, 0.0031]	[-0.0074, -0.0002]
CI Rel. to Pre-Per. Mean	[-1.12%, 0.68%]	[-4.84%, 1.15%]	[-5.55%, 3.32%]	[-0.86%, 6.96%]	[-7.43%, -0.25%]
CI Rel. to Pre-Per. SD	[-1.29%, 0.79%]	[-1.75%, 0.42%]	[-1.43%, 0.86%]	[-0.19%, 1.50%]	[-2.47%, -0.08%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1685760	1685760	1682220	1682220	1682220

**Table B.12:** Heterogeneity by Race: Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0047 (0.0031) {-1.5322}	-0.0047 (0.0031) {-1.5322}	-0.0002 (0.0011) {-0.1444}	-0.0022 (0.0016) {-1.3744}	-0.0001 (0.0002) {-0.4277}
95% CI	[-0.0107, 0.0013]	[-0.0107, 0.0013]	[-0.0022, 0.0019]	[-0.0054, 0.0009]	[-0.0005, 0.0003]
CI Rel. to Pre-Per. Mean	[-2.42%, 0.30%]	[-2.42%, 0.30%]	[-5.62%, 4.85%]	[-6.67%, 1.17%]	[-30.16%, 19.36%]
CI Rel. to Pre-Per. SD	[-2.16%, 0.26%]	[-2.16%, 0.26%]	[-1.15%, 0.99%]	[-1.98%, 0.35%]	[-1.28%, 0.82%]
Treatment $\times$ Post $\times$ Black	0.0032 (0.0047) {0.6669}	0.0032 (0.0047) {0.6669}	-0.0014 (0.0013) {-1.0958}	-0.0017 (0.0018) {-0.9109}	0.0000 (0.0003) {0.0791}
95% CI	[-0.0061, 0.0124]	[-0.0061, 0.0124]	[-0.0039, 0.0011]	[-0.0053, 0.0019]	[-0.0005, 0.0006]
CI Rel. to Pre-Per. Mean	[-1.38%, 2.80%]	[-1.38%, 2.80%]	[-9.79%, 2.77%]	[-6.54%, 2.39%]	[-29.15%, 31.60%]
CI Rel. to Pre-Per. SD	[-1.23%, 2.50%]	[-1.23%, 2.50%]	[-2.00%, 0.57%]	[-1.94%, 0.71%]	[-1.24%, 1.34%]
Treatment $\times$ Post $\times$ Hispanic	-0.0022 (0.0019) {-1.1616}	-0.0022 (0.0019) {-1.1616}	-0.0020 (0.0010) {-1.9964}	0.0004 (0.0012) {0.3107}	-0.0001 (0.0002) {-0.5573}
95% CI	[-0.0059, 0.0015]	[-0.0059, 0.0015]	[-0.0040, -0.0000]	[-0.0019, 0.0027]	[-0.0005, 0.0003]
CI Rel. to Pre-Per. Mean	[-1.33%, 0.34%]	[-1.33%, 0.34%]	[-10.11%, -0.09%]	[-2.39%, 3.30%]	[-28.15%, 15.69%]
CI Rel. to Pre-Per. SD	[-1.19%, 0.30%]	[-1.19%, 0.30%]	[-2.06%, -0.02%]	[-0.71%, 0.98%]	[-1.20%, 0.67%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1685760	1685760	1685760	1685760	1685760

**Table B.13:** Heterogeneity by Race: Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-7.4605 (8.2599) {-0.9032}	-12.8178 (10.6149) {-1.2075}	9.6565 (89.2678) {0.1082}	235.5069 (121.0937) {1.9448}
95% CI	[-23.6500, 8.7289]	[-33.6231, 7.9875]	[-165.3084, 184.6215]	[-1.8368, 472.8506]
CI Rel. to Pre-Per. Mean	[-4.54%, 1.68%]	[-6.17%, 1.46%]	[-2.90%, 3.24%]	[-0.03%, 6.74%]
CI Rel. to Pre-Per. SD	[-3.03%, 1.12%]	[-4.33%, 1.03%]	[-2.35%, 2.63%]	[-0.02%, 4.98%]
Treatment $\times$ Post $\times$ Black	2.2986 (7.4664) {0.3079}	2.6180 (14.5358) {0.1801}	-23.9650 (98.0310) {-0.2445}	-123.9023 (173.9054) {-0.7125}
95% CI	[-12.3356, 16.9328]	[-25.8722, 31.1082]	[-216.1058, 168.1758]	[-464.7569, 216.9524]
CI Rel. to Pre-Per. Mean	[-2.37%, 3.25%]	[-4.74%, 5.70%]	[-3.79%, 2.95%]	[-6.63%, 3.09%]
CI Rel. to Pre-Per. SD	[-1.58%, 2.17%]	[-3.33%, 4.00%]	[-3.08%, 2.39%]	[-4.89%, 2.28%]
Treatment $\times$ Post $\times$ Hispanic	7.5863 (5.9572) {1.2735}	3.0201 (9.7839) {0.3087}	51.2413 (61.2541) {0.8365}	-24.1455 (88.3751) {-0.2732}
95% CI	[-4.0897, 19.2624]	[-16.1563, 22.1965]	[-68.8168, 171.2995]	[-197.3607, 149.0697]
CI Rel. to Pre-Per. Mean	[-0.79%, 3.70%]	[-2.96%, 4.07%]	[-1.21%, 3.00%]	[-2.81%, 2.13%]
CI Rel. to Pre-Per. SD	[-0.52%, 2.47%]	[-2.08%, 2.86%]	[-0.98%, 2.44%]	[-2.08%, 1.57%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	688516	342388	744025	508758

### B.3 Heterogeneity by Officer Age

**Table B.14:** Heterogeneity by Age: Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0079 (0.0036) {-2.2039}	-0.0011 (0.0023) {-0.4771}	-0.0005 (0.0012) {-0.4222}	0.0006 (0.0005) {1.3522}	0.0033 (0.0022) {1.5288}
95% CI	[-0.0150, -0.0009]	[-0.0056, 0.0034]	[-0.0029, 0.0019]	[-0.0003, 0.0016]	[-0.0009, 0.0075]
CI Rel. to Pre-Per. Mean	[-2.62%, -0.15%]	[-4.79%, 2.91%]	[-4.62%, 2.98%]	[-0.66%, 3.59%]	[-0.94%, 7.57%]
CI Rel. to Pre-Per. SD	[-3.03%, -0.18%]	[-1.73%, 1.06%]	[-1.19%, 0.77%]	[-0.14%, 0.77%]	[-0.31%, 2.52%]
Treatment $\times$ Post $\times$ Median Age +	0.0009 (0.0036) {0.2474}	0.0017 (0.0018) {0.9633}	0.0018 (0.0013) {1.4620}	-0.0010 (0.0008) {-1.3091}	-0.0031 (0.0022) {-1.4413}
95% CI	[-0.0062, 0.0080]	[-0.0018, 0.0052]	[-0.0006, 0.0043]	[-0.0025, 0.0005]	[-0.0073, 0.0011]
CI Rel. to Pre-Per. Mean	[-1.08%, 1.39%]	[-1.54%, 4.51%]	[-1.00%, 6.90%]	[-5.71%, 1.14%]	[-7.37%, 1.12%]
CI Rel. to Pre-Per. SD	[-1.25%, 1.61%]	[-0.56%, 1.63%]	[-0.26%, 1.78%]	[-1.23%, 0.24%]	[-2.45%, 0.37%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1685760	1685760	1682220	1682220	1682220

**Table B.15:** Heterogeneity by Age: Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0055 (0.0030) {-1.8498}	-0.0055 (0.0030) {-1.8498}	-0.0016 (0.0011) {-1.5208}	-0.0020 (0.0020) {-0.9992}	-0.0001 (0.0003) {-0.2582}
95% CI	[-0.0113, 0.0003]	[-0.0113, 0.0003]	[-0.0037, 0.0005]	[-0.0060, 0.0020]	[-0.0006, 0.0005]
CI Rel. to Pre-Per. Mean	[-2.55%, 0.07%]	[-2.55%, 0.07%]	[-9.30%, 1.17%]	[-7.45%, 2.42%]	[-34.03%, 26.11%]
CI Rel. to Pre-Per. SD	[-2.28%, 0.07%]	[-2.28%, 0.07%]	[-1.90%, 0.24%]	[-2.21%, 0.72%]	[-1.45%, 1.11%]
Treatment $\times$ Post $\times$ Median Age +	0.0011 (0.0028) {0.3964}	0.0011 (0.0028) {0.3964}	0.0008 (0.0011) {0.6841}	-0.0007 (0.0014) {-0.4767}	-0.0000 (0.0002) {-0.1139}
95% CI	[-0.0044, 0.0066]	[-0.0044, 0.0066]	[-0.0014, 0.0029]	[-0.0035, 0.0021]	[-0.0005, 0.0004]
CI Rel. to Pre-Per. Mean	[-0.99%, 1.49%]	[-0.99%, 1.49%]	[-3.50%, 7.25%]	[-4.28%, 2.60%]	[-26.45%, 23.54%]
CI Rel. to Pre-Per. SD	[-0.88%, 1.33%]	[-0.88%, 1.33%]	[-0.71%, 1.48%]	[-1.27%, 0.77%]	[-1.12%, 1.00%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1685760	1685760	1685760	1685760	1685760



**Table B.16:** Heterogeneity by Age: Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-6.7345 (6.8751) {-0.9795}	-18.7931 (9.9244) {-1.8936}	19.0407 (87.1526) {0.2185}	142.7135 (136.8102) {1.0431}
95% CI	[-20.2098, 6.7407]	[-38.2449, 0.6587]	[-151.7785, 189.8598]	[-125.4346, 410.8615]
CI Rel. to Pre-Per. Mean	[-3.88%, 1.29%]	[-7.01%, 0.12%]	[-2.66%, 3.33%]	[-1.79%, 5.86%]
CI Rel. to Pre-Per. SD	[-2.59%, 0.86%]	[-4.92%, 0.08%]	[-2.16%, 2.70%]	[-1.32%, 4.32%]
Treatment $\times$ Post $\times$ Median Age +	4.2880 (5.5039) {0.7791}	18.3133 (9.6619) {1.8954}	-11.1128 (48.5686) {-0.2288}	107.0541 (87.4869) {1.2237}
95% CI	[-6.4996, 15.0756]	[-0.6239, 37.2506]	[-106.3072, 84.0817]	[-64.4202, 278.5284]
CI Rel. to Pre-Per. Mean	[-1.25%, 2.90%]	[-0.11%, 6.83%]	[-1.86%, 1.47%]	[-0.92%, 3.97%]
CI Rel. to Pre-Per. SD	[-0.83%, 1.93%]	[-0.08%, 4.79%]	[-1.51%, 1.20%]	[-0.68%, 2.93%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	688516	342388	744025	508758

## B.4 Heterogeneity by Officer Experience

**Table B.17:** Heterogeneity by Experience: Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0071 (0.0035) {-2.0224}	-0.0005 (0.0026) {-0.1994}	-0.0014 (0.0013) {-1.0801}	0.0009 (0.0006) {1.6922}	0.0025 (0.0017) {1.4551}
95% CI	[-0.0139, -0.0002]	[-0.0056, 0.0045]	[-0.0041, 0.0012]	[-0.0001, 0.0020]	[-0.0009, 0.0058]
CI Rel. to Pre-Per. Mean	[-2.43%, -0.04%]	[-4.81%, 3.92%]	[-6.50%, 1.88%]	[-0.34%, 4.60%]	[-0.86%, 5.79%]
CI Rel. to Pre-Per. SD	[-2.81%, -0.04%] (0.0002) {-0.1995}	[-1.74%, 1.42%] (0.0001) {0.2429}	[-1.68%, 0.49%] (0.0001) {2.0279}	[-0.07%, 0.99%] (0.0000) {-2.0501}	[-0.28%, 1.92%] (0.0001) {-0.5075}
95% CI	[-0.0005, 0.0004]	[-0.0002, 0.0003]	[0.0000, 0.0004]	[-0.0002, -0.0000]	[-0.0004, 0.0002]
CI Rel. to Pre-Per. Mean	[-0.08%, 0.07%]	[-0.18%, 0.23%]	[0.01%, 0.58%]	[-0.35%, -0.01%]	[-0.35%, 0.21%]
CI Rel. to Pre-Per. SD	[-0.09%, 0.08%]	[-0.07%, 0.08%]	[0.00%, 0.15%]	[-0.08%, -0.00%]	[-0.12%, 0.07%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1685760	1685760	1682220	1682220	1682220

**Table B.18:** Heterogeneity by Experience: Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0053 (0.0032) {-1.6429}	-0.0053 (0.0032) {-1.6429}	-0.0022 (0.0012) {-1.7872}	-0.0026 (0.0021) {-1.2135}	-0.0000 (0.0003) {-0.0119}
95% CI	[-0.0115, 0.0010]	[-0.0115, 0.0010]	[-0.0046, 0.0002]	[-0.0067, 0.0016]	[-0.0006, 0.0006]
CI Rel. to Pre-Per. Mean	[-2.60%, 0.23%]	[-2.60%, 0.23%]	[-11.53%, 0.53%]	[-8.28%, 1.95%]	[-30.97%, 30.59%]
CI Rel. to Pre-Per. SD	[-2.32%, 0.20%]	[-2.32%, 0.20%]	[-2.36%, 0.11%]	[-2.45%, 0.58%]	[-1.32%, 1.30%]
	(0.0002) {0.1516}	(0.0002) {0.1516}	(0.0001) {1.5472}	(0.0001) {0.2281}	(0.0000) {-0.6452}
95% CI	[-0.0004, 0.0005]	[-0.0004, 0.0005]	[-0.0000, 0.0002]	[-0.0001, 0.0002]	[-0.0000, 0.0000]
CI Rel. to Pre-Per. Mean	[-0.09%, 0.10%]	[-0.09%, 0.10%]	[-0.06%, 0.54%]	[-0.17%, 0.21%]	[-1.81%, 0.91%]
CI Rel. to Pre-Per. SD	[-0.08%, 0.09%]	[-0.08%, 0.09%]	[-0.01%, 0.11%]	[-0.05%, 0.06%]	[-0.08%, 0.04%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1685760	1685760	1685760	1685760	1685760

**Table B.19:** Heterogeneity by Experience: Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-10.6184 (8.4208) {-1.2610}	-17.1412 (11.4755) {-1.4937}	36.6466 (89.8438) {0.4079}	128.9146 (152.1530) {0.8473}
95% CI	[-27.1232, 5.8864]	[-39.6332, 5.3508]	[-139.4472, 212.7404]	[-169.3054, 427.1345]
CI Rel. to Pre-Per. Mean	[-5.21%, 1.13%]	[-7.27%, 0.98%]	[-2.44%, 3.73%]	[-2.41%, 6.09%]
CI Rel. to Pre-Per. SD	[-3.48%, 0.75%] (0.4947) {1.4048}	[-5.10%, 0.69%] (0.7738) {1.0134}	[-1.99%, 3.03%] (3.4782) {-0.7470}	[-1.78%, 4.50%] (7.3445) {1.0141}
95% CI	[-0.2746, 1.6645]	[-0.7325, 2.3007]	[-9.4153, 4.2191]	[-6.9468, 21.8434]
CI Rel. to Pre-Per. Mean	[-0.05%, 0.32%]	[-0.13%, 0.42%]	[-0.17%, 0.07%]	[-0.10%, 0.31%]
CI Rel. to Pre-Per. SD	[-0.04%, 0.21%]	[-0.09%, 0.30%]	[-0.13%, 0.06%]	[-0.07%, 0.23%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	688516	342388	744025	508758

## B.5 Heterogeneity by Call Response Volume in 2021

**Table B.20:** Heterogeneity by Call Response Volume in 2021: Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0078 (0.0038) {-2.0578}	0.0001 (0.0018) {0.0615}	0.0016 (0.0014) {1.1785}	0.0004 (0.0007) {0.5673}	0.0011 (0.0024) {0.4565}
95% CI	[-0.0153, -0.0004]	[-0.0034, 0.0036]	[-0.0011, 0.0043]	[-0.0009, 0.0017]	[-0.0036, 0.0058]
CI Rel. to Pre-Per. Mean	[-2.67%, -0.06%]	[-2.95%, 3.14%]	[-1.70%, 6.81%]	[-2.10%, 3.82%]	[-3.63%, 5.84%]
CI Rel. to Pre-Per. SD	[-3.08%, -0.08%]	[-1.07%, 1.14%]	[-0.44%, 1.76%]	[-0.45%, 0.82%]	[-1.21%, 1.94%]
Treatment $\times$ Post $\times$ Med. Call Volume 2021	-0.0006 (0.0034) {-0.1847}	-0.0001 (0.0019) {-0.0612}	-0.0022 (0.0014) {-1.5293}	-0.0002 (0.0010) {-0.2386}	0.0018 (0.0027) {0.6563}
95% CI	[-0.0073, 0.0061]	[-0.0039, 0.0036]	[-0.0051, 0.0006]	[-0.0022, 0.0017]	[-0.0035, 0.0071]
CI Rel. to Pre-Per. Mean	[-1.28%, 1.06%]	[-3.34%, 3.14%]	[-8.07%, 1.00%]	[-4.92%, 3.85%]	[-3.56%, 7.15%]
CI Rel. to Pre-Per. SD	[-1.48%, 1.22%]	[-1.21%, 1.14%]	[-2.09%, 0.26%]	[-1.06%, 0.83%]	[-1.18%, 2.38%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1465240	1465240	1461951	1461951	1461951

**Table B.21:** Heterogeneity by Call Response Volume in 2021: Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0051 (0.0028) {-1.8581}	-0.0051 (0.0028) {-1.8581}	-0.0012 (0.0010) {-1.1360}	-0.0011 (0.0018) {-0.6375}	0.0002 (0.0002) {1.3442}
95% CI	[-0.0105, 0.0003]	[-0.0105, 0.0003]	[-0.0032, 0.0009]	[-0.0047, 0.0024]	[-0.0001, 0.0006]
CI Rel. to Pre-Per. Mean	[-2.37%, 0.06%]	[-2.37%, 0.06%]	[-8.03%, 2.14%]	[-5.78%, 2.94%]	[-6.14%, 32.94%]
CI Rel. to Pre-Per. SD	[-2.11%, 0.06%]	[-2.11%, 0.06%]	[-1.64%, 0.44%]	[-1.71%, 0.87%]	[-0.26%, 1.40%]
Treatment $\times$ Post $\times$ Med. Call Volume 2021	-0.0006 (0.0032) {-0.2026}	-0.0006 (0.0032) {-0.2026}	-0.0008 (0.0010) {-0.7814}	-0.0032 (0.0015) {-2.1349}	-0.0004 (0.0002) {-1.7710}
95% CI	[-0.0068, 0.0055]	[-0.0068, 0.0055]	[-0.0028, 0.0012]	[-0.0062, -0.0003]	[-0.0009, 0.0000]
CI Rel. to Pre-Per. Mean	[-1.54%, 1.25%]	[-1.54%, 1.25%]	[-7.11%, 3.06%]	[-7.70%, -0.33%]	[-49.99%, 2.53%]
CI Rel. to Pre-Per. SD	[-1.37%, 1.12%]	[-1.37%, 1.12%]	[-1.45%, 0.62%]	[-2.28%, -0.10%]	[-2.13%, 0.11%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1465240	1465240	1465240	1465240	1465240

**Table B.22:** Heterogeneity by Call Response Volume in 2021: Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-1.4317 (8.3243) {-0.1720}	-3.5603 (12.2705) {-0.2902}	28.1413 (81.7162) {0.3444}	214.7414 (152.4943) {1.4082}
95% CI	[-17.7473, 14.8840]	[-27.6105, 20.4899]	[-132.0225, 188.3051]	[-84.1474, 513.6302]
CI Rel. to Pre-Per. Mean	[-3.41%, 2.86%]	[-5.06%, 3.76%]	[-2.31%, 3.30%]	[-1.20%, 7.32%]
CI Rel. to Pre-Per. SD	[-2.28%, 1.91%]	[-3.55%, 2.64%]	[-1.88%, 2.68%]	[-0.89%, 5.41%]
Treatment $\times$ Post $\times$ Med. Call Volume 2021	-4.7162 (6.9566) {-0.6780}	-4.6203 (10.0814) {-0.4583}	-37.2623 (72.1669) {-0.5163}	18.0950 (102.4178) {0.1767}
95% CI	[-18.3511, 8.9187]	[-24.3799, 15.1393]	[-178.7093, 104.1847]	[-182.6440, 218.8340]
CI Rel. to Pre-Per. Mean	[-3.53%, 1.71%]	[-4.47%, 2.78%]	[-3.13%, 1.83%]	[-2.60%, 3.12%]
CI Rel. to Pre-Per. SD	[-2.35%, 1.14%]	[-3.14%, 1.95%]	[-2.54%, 1.48%]	[-1.92%, 2.30%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	582657	279512	632757	422887

## B.6 Heterogeneity by Times Treated

**Table B.23:** Heterogeneity by Times Treated: Work Outcomes

Dependent Variable:	Reports to Work	Any Voluntary Overtime Hours	Calls in Sick	Injured on the Job	Paid Time Off
Treatment $\times$ Post	-0.0110 (0.0055) {-1.9956}	-0.0010 (0.0026) {-0.3950}	0.0014 (0.0016) {0.8771}	-0.0002 (0.0008) {-0.2079}	0.0070 (0.0044) {1.5676}
95% CI	[-0.0217, -0.0002]	[-0.0060, 0.0040]	[-0.0018, 0.0047]	[-0.0018, 0.0015]	[-0.0017, 0.0157]
CI Rel. to Pre-Per. Mean	[-3.80%, -0.03%]	[-5.22%, 3.47%]	[-2.85%, 7.46%]	[-4.13%, 3.34%]	[-1.75%, 15.74%]
CI Rel. to Pre-Per. SD	[-4.39%, -0.04%] (0.0058) {0.6989}	[-1.89%, 1.26%] (0.0020) {0.4630}	[-0.74%, 1.93%] (0.0016) {-0.7567}	[-0.89%, 0.72%] (0.0009) {0.4012}	[-0.58%, 5.23%] (0.0042) {-1.4393}
95% CI	[-0.0073, 0.0154]	[-0.0030, 0.0049]	[-0.0043, 0.0019]	[-0.0014, 0.0021]	[-0.0144, 0.0022]
CI Rel. to Pre-Per. Mean	[-1.28%, 2.70%]	[-2.60%, 4.22%]	[-6.81%, 3.01%]	[-3.18%, 4.81%]	[-14.50%, 2.22%]
CI Rel. to Pre-Per. SD	[-1.48%, 3.12%]	[-0.94%, 1.53%]	[-1.76%, 0.78%]	[-0.68%, 1.04%]	[-4.82%, 0.74%]
Pre-Period Mean	0.5720	0.1159	0.0626	0.0443	0.0995
Pre-Period SD	0.4948	0.3201	0.2423	0.2057	0.2993
Clusters	30	30	30	30	30
Observations	1685760	1685760	1682220	1682220	1682220



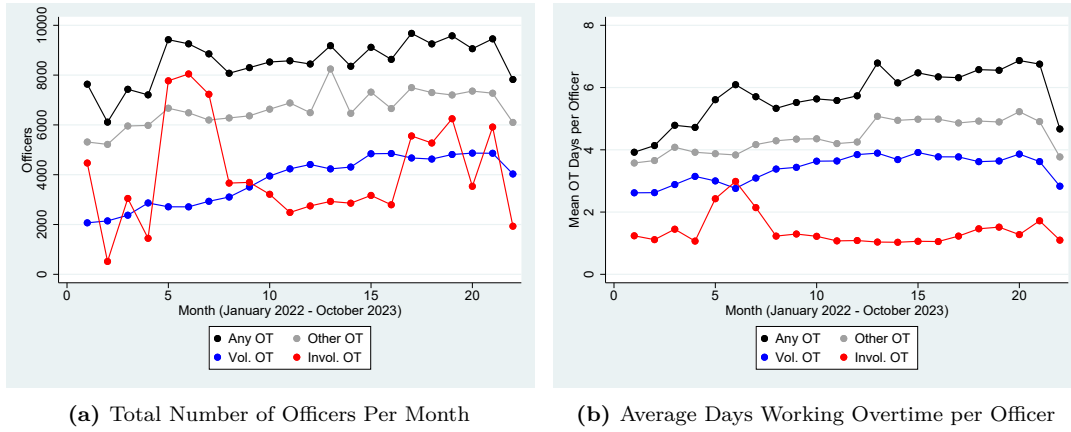
**Table B.24:** Heterogeneity by Times Treated: Enforcement Outcomes

Dependent Variable:	Any Calls for Service	Total Calls for Service	Any Investigatory Stop	Any Arrests	Any Use of Force
Treatment $\times$ Post	-0.0090 (0.0046) {-1.9543}	-0.0090 (0.0046) {-1.9543}	-0.0023 (0.0011) {-2.1283}	-0.0047 (0.0024) {-1.9796}	-0.0006 (0.0004) {-1.3986}
95% CI	[-0.0181, 0.0000]	[-0.0181, 0.0000]	[-0.0044, -0.0002]	[-0.0094, -0.0000]	[-0.0015, 0.0002]
CI Rel. to Pre-Per. Mean	[-4.08%, 0.01%]	[-4.08%, 0.01%]	[-10.96%, -0.45%]	[-11.68%, -0.06%]	[-80.92%, 13.53%]
CI Rel. to Pre-Per. SD	[-3.64%, 0.01%]	[-3.64%, 0.01%]	[-2.24%, -0.09%]	[-3.46%, -0.02%]	[-3.44%, 0.58%]
	(0.0047) {1.0133}	(0.0047) {1.0133}	(0.0012) {1.0120}	(0.0023) {1.1931}	(0.0004) {1.5886}
95% CI	[-0.0045, 0.0140]	[-0.0045, 0.0140]	[-0.0011, 0.0036]	[-0.0018, 0.0073]	[-0.0001, 0.0014]
CI Rel. to Pre-Per. Mean	[-1.01%, 3.16%]	[-1.01%, 3.16%]	[-2.83%, 8.88%]	[-2.19%, 9.02%]	[-7.90%, 75.52%]
CI Rel. to Pre-Per. SD	[-0.90%, 2.82%]	[-0.90%, 2.82%]	[-0.58%, 1.81%]	[-0.65%, 2.67%]	[-0.34%, 3.21%]
Pre-Period Mean	0.4433	0.4433	0.0400	0.0807	0.0018
Pre-Period SD	0.4968	0.4968	0.1961	0.2724	0.0425
Clusters	30	30	30	30	30
Observations	1685760	1685760	1685760	1685760	1685760

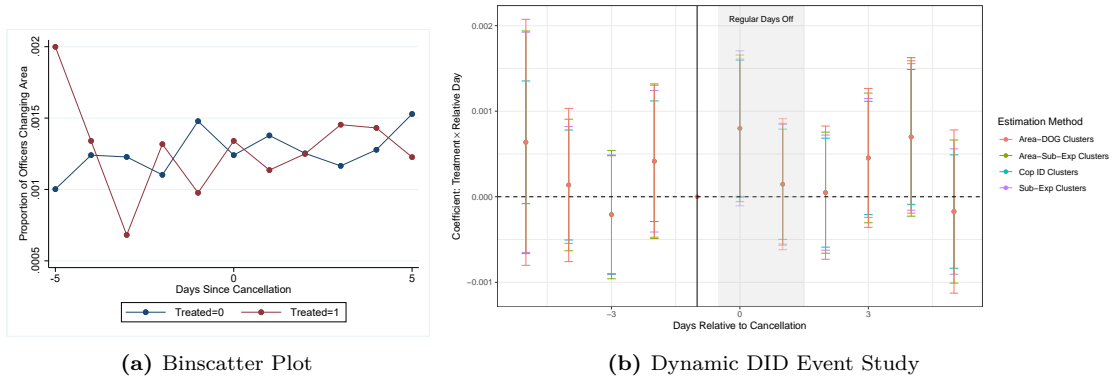
**Table B.25:** Heterogeneity by Times Treated: Response Times

Dependent Variable:	Dispatch to On-Scene	Acknowledge to On-Scene	Dispatch to Cleared	Acknowledge to Cleared
Treatment $\times$ Post	-25.8934 (12.0831) {-2.1429}	-35.1325 (12.3891) {-2.8358}	-14.7456 (121.0284) {-0.1218}	19.4031 (179.6277) {0.1080}
95% CI	[-49.5764, -2.2105]	[-59.4151, -10.8498]	[-251.9613, 222.4701]	[-332.6672, 371.4733]
CI Rel. to Pre-Per. Mean	[-9.52%, -0.42%]	[-10.90%, -1.99%]	[-4.42%, 3.90%]	[-4.74%, 5.30%]
CI Rel. to Pre-Per. SD	[-6.36%, -0.28%] (8.9938) {2.8021}	[-7.64%, -1.40%] (9.3931) {3.1725}	[-3.59%, 3.17%] (88.4530) {0.3875}	[-3.50%, 3.91%] (148.4284) {1.3869}
95% CI	[7.5732, 42.8288]	[11.3888, 48.2099]	[-139.0947, 207.6410]	[-85.0618, 496.7775]
CI Rel. to Pre-Per. Mean	[1.45%, 8.23%]	[2.09%, 8.84%]	[-2.44%, 3.64%]	[-1.21%, 7.08%]
CI Rel. to Pre-Per. SD	[0.97%, 5.49%]	[1.47%, 6.20%]	[-1.98%, 2.96%]	[-0.90%, 5.23%]
Pre-Period Mean	520.5279	545.2795	5705.6480	7012.9797
Pre-Period SD	780.0771	777.3455	7022.0205	9501.0454
Clusters	30	30	30	30
Observations	688516	342388	744025	508758

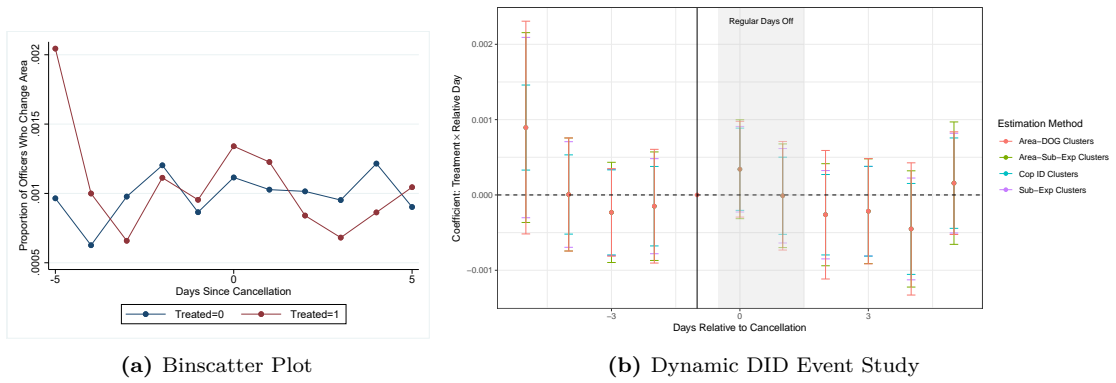
## C Figures



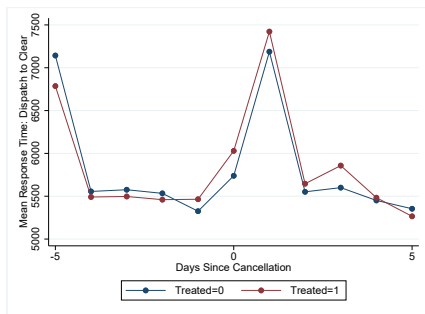
**Figure C.18:** Monthly Overtime by Type, 2022–2023



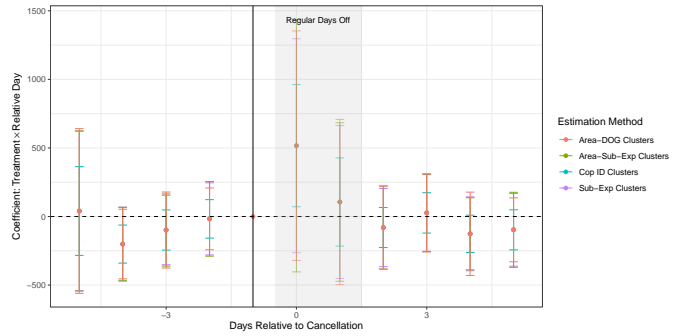
**Figure C.19:** Change in DOG Assignment



**Figure C.20:** Change in Area Assignment

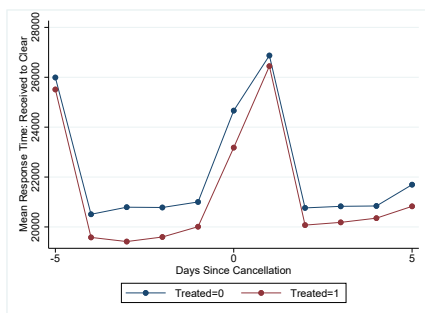


(a) Binscatter Plot

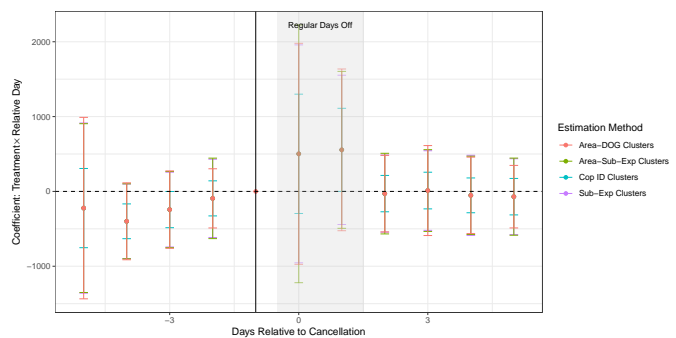


(b) Dynamic DID Event Study

**Figure C.21:** Time from Call Received by Dispatcher to Cleared in Seconds



(a) Binscatter Plot



(b) Dynamic DID Event Study

**Figure C.22:** Time from Officer Dispatched to Cleared