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

Katie Bollman* Ariel Gómez† Matthew Ross‡ CarlyWill Sloan§

26 March 2026

Abstract

Can mandatory overtime effectively serve as a buffer against labor shortages, or does the cost of increased exhaustion and diminished productivity outweigh the benefits? We address this critical question by analyzing a mandatory overtime policy implemented by the Chicago Police Department in 2022-2023. This policy created both cross-sectional and temporal variation in overtime assignments, which we leverage in a "stacked" difference-in-differences design. Using a comprehensive dataset of daily officer-level records, we evaluate the effects of assignment to mandatory overtime on officer behavior, wellness, and on-the-job performance. Our primary findings indicate that being assigned one additional shift does not meaningfully affect officer injuries, sick leave, or overall absence rates. Subsequent attendance declines slightly (approximately 1 percent), while overtime, injuries, and sickness are unchanged and precisely estimated around zero. We find no meaningful changes in calls for service. For enforcement, arrests decline modestly, while use of force is imprecisely estimated. The reduction in arrests is concentrated in less serious, non-index offenses—particularly person offenses—while estimates for serious (index) arrests are near zero. Overall, the evidence is consistent with modest reductions in lower-level discretionary enforcement and no deterioration in officer wellness.

Keywords: Policing, Productivity, Public Safety

JEL: J21, J45, K42

*Oregon State University

†Northeastern University, School of Public Policy and Urban Affairs, Renaissance Park, 310, 1135 Tremont St, Boston, MA 02120, ar.gomez@northeastern.edu

‡Northeastern University

§USMA at West Point. The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Military Academy, the Department of the Army, or the Department of Defense.

1 Introduction

Across the United States, public sector institutions, including schools, hospitals, and police departments, are grappling with severe staffing shortages and challenging work environments (Doan et al., 2024, Maryland Office of Legislative Audits, 2024, Pandemic Response Accountability Committee, 2023). These challenges have been exacerbated by the COVID-19 pandemic and the retirement of an aging workforce. This context is particularly salient for law enforcement agencies, where high vacancy rates have led to concerns about increased workloads and declining officer morale (City of Chicago Office of Inspector General, 2022, Police Executive Research Forum, 2023), further complicated by shifts in public sentiment and increasingly tight budgets (International Association of Chiefs of Police, 2020, National Conference of State Legislatures, 2021). Supported by a growing body of literature, the prevailing view is that excessive workloads negatively affect workers' productivity and wellbeing. However, the effects of modest and strategically distributed increases in workload- common responses for managing staffing shortages in law enforcement- are much less well understood.

Our paper fills this gap in the existing literature by providing direct evidence on how a strategically scheduled form of short-term mandatory overtime affects officer behavior, wellness, and productivity. We study the Chicago Police Department (CPD), which in 2022 introduced a new system of cyclical, pre-scheduled cancellations of Regular Days Off (RDOs) to address staffing shortages and meet minimum staffing requirements. Although CPD had relied increasingly on mandatory overtime since the COVID-19 pandemic, we focus on a distinct policy tied to its tiered deployment model that operated from February 2022 through early 2024. Under this policy, the first of an officer's two scheduled RDOs was canceled once every six weeks according to a predetermined annual scheduling assignment, resulting in a single extended workday followed by a day of rest. Unlike more traditional forms of mandatory overtime, these cancellations were predetermined, not tied to specific events, and affected hundreds of officers at a time. As a result, the policy generates plausibly exogenous variation in short-run workload that is unrelated to individual officer behavior or short-term operational needs.

To estimate the effects of this policy, we use linked personnel, dispatch, and scheduling data from February 2022 through December 2023. Using department-wide "tiered deployment" schedules that specify a matrix of RDO cancellations, we identify cancellation events and construct a balanced panel of officers with rotating work-rest patterns. Our empirical strategy compares officers with otherwise identical work schedules, distinguishing those whose RDO is canceled from those who remain on their usual schedule, and tracks outcomes in the days before and after the cancellation. We implement a stacked event-study design with

officer and date fixed effects specific to each cancellation event, thereby avoiding the early-versus-late comparisons that can bias two-way fixed-effects estimates under staggered rollout (de Chaisemartin and D'Haultfoeuille, 2023, Deshpande and Li, 2019, Goodman-Bacon, 2021, Wing et al., 2023). This design isolates policy-induced short-run variation in workload, allows us to track outcomes with high temporal precision, and provides evidence on an increasingly common strategy for managing chronic staffing shortages in policing and other frontline public-sector settings.

Leveraging our causal identification strategy, we show that assignment to a canceled RDO significantly increases an officer's likelihood of reporting to work that day. In the full analytic sample, untreated officers voluntarily work on their RDOs at a rate of 27 percent, while treated officers work at a rate of 48 percent, implying an 21 percentage point increase in attendance due to the policy. By comparison, approximately 65 percent of officers report to work on a typical scheduled workday, suggesting substantial compliance with the policy, albeit at an attendance rate below that observed on a regular workday. Given that compliance is high but not perfect, we adopt an intent-to-treat framework that accounts for the possibility that realized compliance with the policy is endogenous.

We next examine how assignment to work on an RDO affects outcomes observed after the canceled RDO by comparing treatment and control officers on days in the subsequent workweek when both are scheduled to work their ordinary shifts. Looking first at measures of wellness, we find little evidence that the policy meaningfully affects officer attendance or health once officers return to work. Assignment to a canceled RDO reduces the probability that an officer reports to work on subsequent days by just 0.66 percentage points, or about 1 percent relative to the baseline rate. We find similarly small effects on officer wellness. Injury-related absences are essentially unchanged, with point estimates near zero and confidence intervals that rule out increases of more than about 2 percent. Sick leave rises slightly, by about 0.45 percent relative to the baseline rate, and the confidence intervals allow us to rule out increases of more than approximately 2 percent. Taken together, these estimates suggest that this form of short-term mandatory overtime has little meaningful effect on subsequent attendance or short-run measures of health.

Turning next to measures of productivity and enforcement, we again find little evidence of meaningful effects. Calls for service decline by about 1 percent, and investigatory stops decline by about 3.5 percent. The estimate for calls for service is precise and the event-study exhibits clean pre-trends, suggesting little change in how officers respond to calls. In contrast, the estimates for investigatory stops are noisier, with somewhat imperfect pre-trends, so we interpret these results more cautiously. For post-contact outcomes, we estimate a small but statistically significant decline in arrests of about 2.4 percent, while use of force declines by

an imprecise 2 percent. Decomposing arrests by charge, we find that the underlying effect for index crimes is close to zero, and the confidence intervals rule out decreases larger than roughly 3.5 percent. Instead, the decline in arrests appears to be driven by reductions in less serious, non-index offenses (a decline of about 9 percent), particularly person offenses, along with smaller declines in property and public-order offenses. Taken together, this pattern suggests a modest reduction in low-level post-contact activity rather than a consequential change in how officers interact with citizens following their exposure to the additional mandatory overtime shift.

Across our outcome measures, we find little evidence of differential pre-trends, and the results are generally robust to alternative specifications, including different fixed-effects structures and longer event-study windows. A likely explanation for these modest effects is that the mandatory overtime shifts induced by the policy do not resemble ordinary patrol work. Descriptive evidence from the data shows that activity levels on policy-induced mandatory overtime shifts are substantially lower than on regular workdays: officers are much less likely to respond to calls for service, make stops, or make arrests on these shifts than during standard patrol. At the same time, activity on mandatory overtime shifts is broadly similar to that observed when officers work voluntary overtime shifts. Administrative overtime records reinforce this interpretation, showing that nearly all RDO overtime was classified as "Special Event" work, much of it tied to strategic deployment initiatives designed to increase visible police presence rather than respond to calls for service. In other words, the policy appears primarily to assign officers to the same types of lower-intensity overtime duties that, absent the policy, they would more often select into voluntarily. Taken together, these patterns help explain the modest effects we estimate and suggest that our findings are informative about police overtime more broadly because overtime assignments often involve lower-intensity, visibility-oriented duties rather than ordinary patrol work.

This study contributes to three strands of literature: burnout, overtime and police performance, and public safety management. The existing literature on high-burnout occupations documents the long-term consequences of cumulative fatigue and extreme stressors on worker productivity and turnover (Blackburn et al., 2023, Cho et al., 2024). Research on healthcare workers and other frontline occupations consistently shows that long hours and stressful conditions can reduce productivity, impair judgment, and increase errors (Blackburn et al., 2023, Lu and Lu, 2017, Nekoei et al., 2024, Sachiko and Isamu, 2016). A related literature focuses on short-run increases in hours or consecutive days worked, particularly in policing where these policies are prevalent. Descriptive evidence from the NYPD finds that overtime correlates with higher use of force, more complaints, and officer injuries (New York City Department of Investigation, 2023a). While burnout reflects chronic strain, even

temporary workload increases can mimic early symptoms of burnout and shape officer behavior. Complementary evidence comes from experimental settings: fatigued night-shift officers perform worse in simulated scenarios (James, 2018).

Our project is most closely related to concurrent work by Ferrazares (2025), who also uses data from the Chicago Police Department. Their study leverages a well-accepted identification strategy that exploits idiosyncratic variation in scheduling assignments (Ba et al., 2021, Gudgeon et al., 2024, Holz et al., 2023). Rather than examining a discrete policy, Ferrazares investigates the cumulative effects of consecutive workdays, finding that each additional day worked increases the risk of both use of force and officer injury— a pattern consistent with fatigue and burnout. Our study differs in two important ways. First, we focus on a specific scheduling policy that imposes a single mandatory overtime day on officers and is followed by a guaranteed period of rest. Second, by focusing on discrete policy-induced scheduling shocks rather than cumulative exposure, our design identifies the immediate, short-run effects of overtime as a management tool rather than its long-run consequences.

As Owens and Ba (2021) emphasize, police departments must weigh the crime-reducing benefits of additional officer activity against its costs, a tradeoff that becomes especially relevant when agencies rely on overtime to manage staffing shortages. Our findings speak directly to this tradeoff and align with a strain of the management literature that identifies contexts in which marginal increases in workload can be absorbed without measurable harm. In healthcare, Kuntz et al. (2015) document a "tipping point" in hospital mortality, showing that hospitals can absorb marginal patients until very high occupancy rates, while Kc and Terwiesch (2009) find that hospitals can increase service rates under rising workloads. Comparable evidence in policing comes from Cho et al. (2024), who show that enforcement activity falls after a line-of-duty death without increasing crime, suggesting that adjustments in policing intensity can be absorbed without jeopardizing public safety. Taken together, the existing literature on the cumulative effects of burnout and this strain of management literature suggests that fatigue-related harms may arise over the long run but only when exposure is sustained without adequate recovery. Our study contributes to this literature by analyzing a distinct scheduling margin— a single mandatory overtime day followed by rest— and finding no impact on short-term productivity and well being.

Finally, a separate and related literature highlights mechanisms that may mitigate the negative impact of excessive work and interventions that may reduce their deleterious effects. For example, Chalfin and Goncalves (2020) demonstrate that police officers' pro-social motivations may influence how they respond to additional demands which can mediate the relationship between workload intensity and outcomes. Several studies suggest that interventions such as peer support and compressed work schedules have the potential to

mitigate burnout. Linos et al. (2022) show that dispatcher peer-support programs reduce burnout, and Fan et al. (2025) find a significant reduction in burnout from a four-day work week. In the short run, our study suggests that these interventions may not be necessary if mandatory overtime is strategically scheduled such that the affected day is followed by a guaranteed period of rest.

Our findings have practical implications for police leaders navigating chronic staffing shortages. While public debate often focuses on sweeping reforms or major recruitment drives, departments typically manage understaffing through marginal adjustments such as overtime and extended shifts. The Chicago Police Department's systematic and cyclical day-off cancellations exemplify such incremental responses, implemented without the confounding influence of emergent operational needs. This setting reflects a broader reality across large police departments, where overtime has become a routine staffing tool and the policy debate increasingly centers on how to govern its use rather than whether to rely on it at all. For example, recent audits of the Baltimore Police Department documented weaknesses in monitoring and approving overtime hours, prompting reforms aimed at strengthening oversight (Baltimore Police Department, 2023, Maryland Office of Legislative Audits, 2023), while the New York City Department of Investigation similarly reports that the NYPD lacks effective safeguards to limit fatigue risks associated with extensive overtime use (New York City Department of Investigation, 2023b). In San Francisco, persistent staffing shortages have led city leaders to expand overtime budgets while simultaneously debating stronger oversight and alternative staffing strategies (San Francisco Chronicle, 2024a,b). Against this backdrop, our results suggest that modest increases in mandatory overtime do not generate measurable short-run harms, and that equally modest reductions may not yield substantial improvements. Our findings may also extend to other public-sector workforces where overtime is a routine stopgap, particularly frontline workers.

2 Institutional Context, Data, and Sample Setup

2.1 Institutional Context

Like many jurisdictions in the United States, CPD experienced gradual declines in the number of sworn officers over the last two decades. These declines became especially acute between 2019 and 2022, when retirements and resignations among sworn Chicago police officers nearly doubled, producing a 10% reduction in patrol officers. Although it is difficult to pinpoint the specific causes of these departures, they coincided with the upheaval of the COVID-19 pandemic and a wave of nationwide protests against excessive use of force by police. In

response to these staffing shortages and the protests that followed the killing of George Floyd, CPD began canceling scheduled RDO in May 2020.

The goal was to ensure sufficient staffing during periods of heightened crime and disorder. Over the subsequent three years, CPD continued to rely on RDO cancellations, effectively mandating overtime across the force. In February 2022, the department formalized this practice by instituting pre-determined cancellations that cycled through groups of officers. Importantly, officers had little choice over when these mandatory days off occurred.[1] Concerns about the toll on staff soon surfaced. A 2022 audit revealed that some officers worked as many as 12 consecutive days with shifts as long as 12 hours in April and May of that year (City of Chicago Office of Inspector General, 2022). The policy remained in place until October 2023, when newly appointed Superintendent Larry Snelling announced a general prohibition on mandatory overtime, limiting it only to holidays and special events beginning in 2024 (Sisk, 2023). While this termination marked the end of the scheduling mandate, it was also driven by a re-evaluation of how these additional hours were being utilized in practice.

Additional evidence from administrative records helps clarify the nature of work performed on RDO overtime days. Table 1 shows that officer activity during both voluntary and involuntary RDO overtime shifts is substantially lower than during normal patrol shifts, with markedly fewer calls for service, stops, and arrests. At the same time, voluntary and involuntary RDO shifts look more similar across these dimensions, suggesting that the assignments imposed by the policy closely resemble the types of overtime shifts officers commonly work voluntarily rather than unusually demanding deployments.[2] Administrative overtime records reinforce this interpretation: nearly all of RDO overtime is categorized as "Special Event" work, much (58%) of which corresponds to strategic deployment initiatives designed to increase visible police presence rather than respond to calls for service.[3]

Consistent with this pattern, a notable portion of mandatory overtime hours were allocated to a high-visibility enforcement strategy colloquially known as "Operation Scarecrow." Officially aimed at deterring crime through static presence, this initiative required officers to man fixed posts in marked vehicles with emergency lights activated, particularly within downtown business districts (Masterson, 2024). The scale of this deployment was substantial, but the nature of the work differed from standard patrol duties: officers on these assignments

[1]Additional mandatory overtime was also assigned during special events and holidays independently of this policy.

[2]Extended overtime denotes time worked beyond an officer's scheduled shift. Consistent with the summary statistics in Table 1, such extensions are typically driven by active enforcement or required follow-up tasks, such as arrest processing.

[3]The overtime data provide relatively clear descriptions for voluntary and involuntary RDO shifts but are less informative for other categories. Some classifications are ambiguous, limiting our ability to fully characterize all overtime assignments.

primarily served as visible deterrents rather than actively responding to calls or engaging in enforcement. The initiative drew criticism for treating officers as "props," a concern cited by Superintendent Snelling when he dismantled the program in late 2023 to prioritize morale and effective deployment (NBC Chicago, 2023, Schuba and Main, 2024). Taken together, the evidence suggests that a meaningful share of mandatory RDO assignments during our study period involved lower-intensity, pre-planned deterrence rather than the high-stress demands of typical patrol work.

The aggregate impact of these initiatives is readily apparent in the raw administrative data. As shown in Figure 1, the cumulative number of days when officers claimed overtime grew by 59% from a mean of 175 thousand between 2014 and 2019 to nearly 250 thousand in 2023. Similarly, the cumulative number of overtime hours grew by 56% over the same period. Between 2019 and 2020, involuntary overtime days and hours increased by 172% and 134%, respectively. The number of both days and hours of involuntary overtime remained at elevated levels through 2023. From 2021 to 2023, the rate of voluntary overtime in terms of days and hours grew by more than a factor of four. These figures highlight the sheer scale of overtime and CPD's growing reliance on both voluntary and involuntary overtime, a broader context that extends beyond our study's focus on a specific mandatory policy.

Understanding CPD's work assignment system is critical for interpreting the overtime policy. Beginning long before the implementation of the mandatory overtime policy, officers at CPD have been assigned to a "day-off group" (DOG) that follows a repeating six-day cycle. For patrol officers, this would typically consist of four consecutive work days followed by two consecutive rest days. DOG assignments go into effect in January of each year and are determined through a seniority-based bidding process in the preceding fall. Although officers can and do change DOGs throughout the year— we observe switching in nearly 40% of officer-years— these changes remain subject to managerial discretion. In addition to a DOG assignment, patrol officers at CPD are also assigned to a particular geographic unit. The city is divided into five patrol areas, which contain 22 districts and 277 beats. While other empirical studies using CPD data have leveraged idiosyncratic variation from January DOG assignments, our strategy uses a distinct source of identifying variation.[4] Specifically, our empirical design relies on the tiered-deployment RDO cancellation matrix, which varies over time by both assigned DOG and patrol area.

As mentioned, the mandatory overtime from the tiered-deployment policy was operationalized through a "cancellation matrix"—a set of pre-determined calendars that revoked RDOs for all officers assigned to a specific combination of DOG and patrol area. Each cycle began

[4]Recent research has also used CPD's DOG system to study how officer race, gender, and experience shape performance (Ba et al., 2021, Gudgeon et al., 2024).

in a particular area, with cancellations cycling sequentially through each DOG assignment within that area. Subsequent cycles repeated this process in a different area. This rotating system meant that the timing of mandatory overtime was driven by a departmental schedule, not by officer choice. For there to be endogenous selection in our setting, one would have to imagine that officers obtain manager approval to move between DOG assignments and assigned areas based on the timing of the canceled RDO on the matrix. We believe, and our event study estimates corroborate, that this is unrealistic and that the policy created plausibly exogenous variation in which officers were required to work on their RDO.

2.2 Data

Our analysis utilizes administrative data primarily obtained from Freedom of Information Act (FOIA) requests filed with the CPD. Officer attendance data is sourced from the CPD's Attendance and Assignment (A&A) database, which is effectively "punch card" data that records the start and end of each shift and includes a detailed paycode. With nearly a hundred options, these paycodes are highly specific, documenting whether an officer was on various types of regular or overtime duty. The data also capture the officer's shift, unit, and beat for a given day, and notes the reason for any absence, such as paid time off, a workplace injury, or an RDO. While the A&A database tracks daily work and absences, we separately use a formal assignments database to observe each officer's official DOG and area assignment over time. We measure treatment by combining these formal assignments with the RDO cancellation matrix. An officer is considered treated by the policy if their DOG-area combination is scheduled to have its RDO canceled. Since supervisors can grant exemptions, resulting in imperfect compliance, we focus on intent-to-treat effects. We define an officer as treated if they are scheduled to work on their RDO, regardless of whether they ultimately do.

We study outcomes in three domains: officer wellness, productivity, and enforcement. Each of these measures captures a distinct channel through which mandatory overtime might affect officers and the public. We measure wellness using various reasons for an absence reported in the A&A data. Specifically, we construct daily indicators for whether an officer (1) was absent due to an injury sustained on the job, or (2) called in sick to work. We also consider attendance and overtime in the week following a canceled RDO as complementary measures of wellness. Together, these outcomes capture whether mandatory overtime increases short-term illness or injury risk as well as whether officers make their own workload adjustments to mitigate the immediate effects of the policy.

We measure productivity by counting the number of incidents involving each officer, either through responding to a call for service (CFS) or by initiating an investigatory stop.

CFS data capture incidents for which a 911 dispatcher requested and received a police response, and thus reflect responses to public demand for police assistance. Investigatory stop data capture officer-initiated encounters in which an officer temporarily detains an individual suspected of criminal activity. We obtained universe files for both incident types from CPD and construct daily officer-level counts of CFS responses and investigatory stops.[5] Since both activities involve some degree of discretion (less in CFS, more in investigatory stops), mandatory overtime may reduce engagement if overburdened officers choose to shirk.

Our third domain— enforcement activity—is closely related to productivity but conceptually distinct. Whereas our productivity measures capture selection into an incident, our enforcement measures capture the intensity of those interactions via the decision to make an arrest or use force . The arrest data provide the universe of CPD arrests with information on charges and all officers involved (primary, secondary, or assisting). Because arrests are relatively rare at the officer-day level, we define an arrest indicator equal to one if an officer contributed to any arrest that day. We further disaggregate arrests into several categories based on FBI offense codes. Specifically, we distinguish between index crimes (serious offenses), traffic violations, and three types of non-index offenses: order-related offenses (e.g., disorderly conduct, drug and liquor violations), person offenses (e.g., simple assault and battery), and property offenses (e.g., fraud and vandalism). We also identify instances in which individuals are released without charges. This categorization allows us to examine whether effects differ across types of enforcement activity that vary in severity, discretion, and required officer effort. Use of force data identifies officers who reported using physical force. CPD requires officers to document force used against "actively resisting" civilians, resulting in reported injury, or involving civilians who threatened, obstructed, or assaulted an officer (Chicago Police Department, 2023).[6]

2.3 Sample Setup

While the policy was active from February 2022 to December 2023, we focus our analysis on the period between May 31, 2022, and October 20, 2023, excluding a three month period between January 3 and April 4, 2023. We limit our study to this specific time frame because the cyclic RDO cancellations were most consistently documented and implemented during these periods.[7] Of the total 13,601 officers (93% of all officers) in the roster that we are

[5]We top code officer daily CFS volume at the 99th percentile or 21 calls in a single day.

[6]In 2022–23, 79% of reports indicated use of control tactics (e.g., holds, restraints); 61.7% reported bodily force without weapons (e.g., punches, tackles); and 4.4% reported use of a weapon (e.g., tasers, batons, firearms).

[7]We omit the first and last months of the policy because of irregular and inconsistent implementation. We omit much of Q1 2023 because of a gap in the availability of cancellation matrices during this period.

able to link across all of our data, we restrict the sample to 7,927 patrol officers because their schedules are most directly impacted by the policy. The majority of patrol officers were assigned to one of six DOGs (61–66). All of these DOG schedules followed a four-days-on to two-days-off rotation. These DOGs are sequentially offset by one day, ensuring consistent staffing throughout the week. We also restrict the sample to patrol officers assigned to the five main geographic areas (Areas 1–5). Citywide units, which operate across multiple areas, were excluded because they did not follow the same cancellation pattern.[8]

Cancellations fall on the first rest day of each DOG cycle. Complying officers therefore worked on that day, rested the next, and then restarted their DOG cycle. We use this structure to construct a series of "sub-experiments". For each cancellation, we compare officers in the same DOG but assigned to different areas such that only some experienced the RDO cancellation. Our analytical sample follows officers for five days before and after each RDO cancellation, capturing the work cycles directly surrounding treatment. This window allows us to measure immediate changes in wellness (injuries, sick leave, attendance), productivity (calls for service and investigatory stops), and enforcement (arrests and use of force). For each sub-experiment, we construct balanced panels, excluding officers we do not observe in all of the data within the respective period. For control units, we additionally require that officers were not treated within an eleven-day window around the focal cancellation.[9]

Pooling across 288 observed cancellations yields a stacked sample covering roughly 48% of patrol officer-dates during the study period.[10] In the analytical sample, 34% of officer-dates fall are treated units and 64% are control units. As shown in Table 2, absences are somewhat common (on an average day, 4.4% of officers call out injured, and 6.1% sick, with additional officers taking paid time off. Officers average 2.4 responses to calls for service per day, initiate 0.06 investigatory stops, make 0.08 arrests, and use force on 0.17% of days. These outcomes provide the empirical foundation for assessing how mandatory overtime reshaped wellness, productivity, and enforcement.

[8]To minimize potential noise from short-term assignment changes, we impute DOG and area by officers' modal assignment within each month.

[9]In particular, we omit officers treated in the work cycle (6 days) prior to the focal cancellation, or during the remainder of the work cycle beginning with the focal RDO (5 days). We provide results from longer- and alternative short-run windows in the Appendix.

[10]The full linked data cover 7,927 officers and 3.2 million officer-days. The stacked sample consists of 5,864 officers and 1.9 million officer-days. Demographics and baseline outcomes are nearly identical between the full and stacked samples, aside from a slightly higher share of Black officers (1.7 percentage points). Detailed balance tables are in Table 2.

3 Empirical Design

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. In effect, our design compares officers in the same DOG who would otherwise have identical schedules, but differ due to RDO cancellations in some areas. As described in Section 2.2, we consider each RDO cancellation as a single natural experiment induced by the overtime policy. By pooling data across sub-experiments (i.e. individual RDO cancellation events) in a stacked sample, our 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. This design avoids problematic comparisons between early and later treated observations that can arise in a two-way fixed-effects specification in the presence of treatment heterogeneity (de Chaisemartin and D'Haultfoeuille, 2023).

To examine the temporal dynamics of our outcomes and assess pre-treatment trends across groups, we estimate models of the form:

$$Y_{iet} = \sum_{t} \delta_t \, D_{ie} \times T_t + \phi_{i \times e} + \lambda_{t \times e} + \epsilon_{iet} \tag{1}$$

where Y_{iet} denotes an outcome of interest for officer i in sub-experiment e at event time t and ϵ_{iet} capture idosyncratic errors. We include officer-by-experiment fixed effects ($\phi_{i \times e}$), which absorb all time-invariant differences in officer behavior within each experiment.[11] Allowing the officer fixed effects to vary by experiment is important because the composition of the sample and the mapping of officers into risk sets can differ across stacks, so an officer's baseline behavior need not be constant across experiments. We also include date-by-experiment fixed effects ($\lambda_{t \times e}$), which capture shocks common to all officers within a given experiment on a given date. Allowing the date fixed effects to vary by experiment accounts for the fact that each stacked experiment is centered on a different event date.

The variable D_{ie} is an indicator equal to one if an officer is scheduled to have their RDO canceled (treatment) during the sub-experiment and zero if they are not (control). The set of variables T_t are indicators capturing event time relative to the focal cancellation. In a model with officer-by-experiment and date-by-experiment fixed effects, the coefficient δ_t on the interaction term $D_{ie} \times T_t$ can be interpreted as a dynamic difference-in-differences estimator

[11] The officer-by-experiment fixed effects subsume lower-level assignment fixed effects, such as area-by-DOG fixed effects. We verify in subsequent analyses that our results are robust to alternative, more granular fixed-effect specifications, including those that more directly account for assignment-level heterogeneity.

capturing the causal effect of the policy. In many specifications, we replace the full set of relative event-time indicators with a single post-cancellation indicator in order to estimate a static difference-in-differences model of the average effect in the days following a cancellation. Although treatment assignment is plausibly exogenous, compliance is imperfect, so these estimates are most appropriately interpreted as intent-to-treat effects.[12] Since treatment assignment is determined by the combination of an officer's area and DOG assignments, we cluster standard errors at the area-by-DOG level.[13]

3.1 Identifying Assumptions

Our empirical strategy enables us to avoid key sources of endogeneity in the relationship between overtime and worker performance. First, because overtime is normally voluntary, individuals who work extra hours might systematically differ from those who do not. We overcome this selection bias by examining mandatory overtime imposed through RDO cancellations; our intent-to-treat estimates are robust to imperfect compliance. Second, the demand for police overtime often correlates with spikes in crime, potentially confounding results. We mitigate this by leveraging the CPD's predetermined, sequential RDO cancellation pattern, which is determined months in advance and orthogonal to immediate crime trends or demand fluctuations. Third, we use a daily officer panel within a difference-in-differences framework, incorporating officer, area - by - date, and DOG- by - date fixed-effects, to account for non-random officer selection into specific assignments, areas, and initial level differences in working conditions. We previously noted that officers deviate from their January assignments and may endogenously select into their DOG or area assignments. However, our data show no evidence of officers being selected within their DOG by area assignments in response to the timing of the RDO cancellation schedule.

Our difference-in-differences identification strategy relies on several standard assumptions. The "no anticipation effects" assumption requires that officers do not change behavior in advance of a cancellation. Empirically, our event study plot in Figure 2b shows no evidence of anticipation effects for absences. In Appendix Figure C.1, we also show that officers do not differentially change DOG or area to evade treatment. Next, the parallel trends assumption requires that treatment and control group outcomes would evolve similarly without the

[12]While it is possible to estimate 2SLS versions of the difference-in-differences specification, we feel that the reduced-form intent-to-treat estimates are more appropriate because the resulting LATE from the IV is difficult to interpret when compliance may itself reflect endogenous responses to the policy. Moreover, the exclusion restriction may fail if assignment affects behavior through channels other than the additional hours worked.

[13]Alternative standard error assumptions like clustering on officer or twoway clustering by DOG and area yield qualitatively similar results.

canceled RDO. This is plausible given our observation of comparable pre-treatment trends for all outcomes; we present event study estimates for the four regular work days preceding treatment for each outcome in section 4. While we are powered to detect minor variations in pre-treatment event study estimates, these statistical violations are best characterized as idiosyncratic or small and do not amount to meaningful evidence of differential pre-trends that would invalidate our inference.

Furthermore, because CPD implemented cancellations sequentially and across the entire city, it is unlikely that our results are driven by localized shocks or unexpected events that happened to overlap with treatment. An alternative concern is that the broad scope of officers affected by cancellations allows our control groups to include officers who were treated in other sub-experiments. However, this does not introduce bias if treatment effects are short-lived – specifically, if they do not persist into subsequent work cycles after our evaluation window. This is reasonable in our institutional setting, where mandatory overtime occurs no more than once a month and each work cycle with a canceled RDO is followed by two days off. In addition, our regressions explicitly control for the number of days since an officer was last treated. As a robustness check, we relax this assumption and allow treatment effects to last for two full work cycles instead of one. As discussed later, the results are highly similar under this alternative specification.

4 Results

4.1 Treatment Compliance

A key requirement for our empirical strategy is that RDO cancellations meaningfully increase time worked, so we begin by examining treatment compliance. Because officers may switch assignments or take alternative leaves throughout the year, we neither expect nor observe perfect attendance (compliance with the policy) for the treated group or perfect absence for the control groups. [14] However, we can clearly demonstrate the validity of the first stage for producing intent-to-treat estimates: on average, officers do work more when their RDO is canceled.

We begin by establishing compliance directly in the raw data. Based on CPD scheduling rules, officers work four consecutive days followed by two regular days off (RDOs), corresponding to event times $t = 0$ and $t = 1$. Figure 2a shows that treatment and control officers have nearly identical attendance on scheduled workdays ($t = -4, \ldots, -1$ and $t = 2, \ldots, 5$),

[14] Importantly, as we describe in section 3 and show in Figure C.1, assignment switching is not observed in anticipation of a canceled RDO.

and both groups experience a large decline in attendance on the scheduled RDO ($t = 1$). On the canceled RDO ($t = 0$), however, treated officers are substantially more likely to report to work, while control officers exhibit the expected decline. In the full sample, about half of treated officers report to work on the cancellation date, compared to much lower attendance among controls. Relative to baseline attendance of approximately 0.65 on regular workdays, this implies that roughly three-quarters of expected work activity is preserved under treatment.

This pattern is confirmed in the event-study estimates. Assignment to treatment increases the probability of reporting to work on the cancellation date by approximately 22 percentage points, or 34 percent relative to the pre-period mean, and the estimate is statistically significant. Importantly, attendance on scheduled workdays is well below 100 percent due to furloughs, vacation, illness, and other administrative absences, so full compliance is not a relevant benchmark. Instead, the appropriate comparison is to typical attendance on regular workdays, making the observed increase economically meaningful.

Turning to the intensive margin, Figure 2c and Figure 2d plot overtime hours worked. On canceled RDOs, treated officers work approximately 2 additional hours relative to the control group, and approximately 3 additional hours in the larger first-stage sample.[15] This increase is consistent with a standard 9-hour shift scaled by the observed compliance rate, implying that officers who comply with the policy are working a full shift on their canceled RDO. Taken together, these results provide strong evidence that a canceled RDO induces the intended behavioral response: officers assigned to treatment work on days they would otherwise have been off.

4.2 Attendance and Officer Wellness

A primary criticism of mandatory overtime, and a key concern motivating CPD to discontinue the policy, is that it may undermine officer wellness. To assess this possibility, we examine whether canceled RDOs affect four outcomes reported in Panel A of Table 3: reports to work, overtime hours, on-the-job injury, and calling in sick.

The raw data and event-study figures show little evidence of differential pre-trends or post-treatment adjustments. As shown in Figure 2a and Figure 2b, treatment and control officers exhibit parallel trends in attendance prior to the canceled RDO and do not meaningfully diverge on subsequent workdays. Similarly, Figure 2c and Figure 2d show no evidence of changes in overtime hours in the post-period. For wellness outcomes, Figure 3a and Figure 3b indicate that injury rates are stable over time, while Figure 3c and Figure 3d

[15] Most overtime in the pre-period likely reflects extended shifts, when officers remain on duty for a few additional hours to complete arrests or other tasks.

show that sick leave fluctuates but follows parallel trends across treatment and control groups both before and after the policy.

Consistent with these visual patterns, the estimates in Panel A of Table 3 are small in magnitude. First, reporting to work declines by −0.00658, or one percent relative to a pre-period mean of 0.6559. In a typical sub-experiment with 623 officers, this corresponds to approximately 4 fewer officers reporting to work. Second, overtime hours decline by −0.00836 relative to a pre-period mean of 0.5513 (a 1.5 percent reduction).

Third, we find no evidence that canceled RDOs increase on-the-job injuries. The estimated effect is 0.00017, essentially zero relative to a pre-period mean of 0.0419. The upper end of the 95 percent confidence interval corresponds to an increase of only about 2 percent. Fourth, we also find no evidence of a meaningful increase in sickness-related absences. The estimate is 0.00029 (0.45 percent), and the upper end of the 95 percent confidence interval corresponds to an increase of only about 2.4 percent. Taken together, the absence of pre-trends, the stability of outcomes in the post-period, and the small and precisely estimated coefficients indicate that canceled RDOs do not meaningfully undermine officer attendance or wellness.

4.3 Officer Productivity and Enforcement

Although officers continue to show up for work following a canceled RDO, mandatory overtime may still affect how productive they are while on patrol. To examine this, we study two central measures of patrol productivity: calls for service (CFS), which capture responses to demand, and investigatory stops, which capture discretionary activity.

The binscatter and event-study plots in Figure 4 show that treatment and control officers exhibit broadly parallel pre-trends, with no evidence of divergence prior to treatment and no meaningful post-treatment dynamics for calls. Pre-trends for investigatory stops are somewhat noisier and not perfectly flat, so we interpret those results more cautiously, but they do not suggest large underlying differences. Consistent with these patterns, Panel C of Table 3 shows small and statistically insignificant effects. Calls for service decline by 0.025 (0.87 percent), while investigatory stops decline by 0.00262 (3.54 percent). The confidence intervals are relatively tight: for calls for service, the upper end of the 95 percent confidence interval implies we can rule out increases larger than about 0.2 percent, and the lower end implies we can rule out declines larger than about 2 percent. For stops, we can rule out increases larger than about 1 percent and declines larger than 8 percent.

We next examine enforcement outcomes, which capture more consequential margins of officer behavior. Panel D of Table 3 shows that the probability of making any arrest

declines by 0.00220 (2.4 percent), while use of force declines by -0.00004 (1.87 percent). In Figure 5, the event study plots show parallel pre-trends for both outcomes. For arrests, the post-treatment estimates are stable and precisely estimated. In contrast, the use of force series is visibly noisier, with greater fluctuation across event time and no clear or systematic pattern following treatment. Importantly, we can rule out very small increases in arrests (0.35 percent). For use of force, however, the outcome is rare and the estimates are correspondingly imprecise. While the 95 percent confidence interval allows for increases of up to approximately 16 percent and declines of 20 percent, the absence of a clear pattern in the event study and the overall noisiness of the series suggest that we do not interpret this as evidence of a meaningful effect.

To better understand the arrest pattern, we decompose arrests by type in Table 4. Index arrests correspond to serious offenses (such as violent and major property crimes), while non-index arrests capture less serious offenses and are further divided into property, person, and public order offenses. We also separately identify arrests that do not result in charges at the police station, although these are rare. This decomposition shows no meaningful decrease in serious (index) arrests. The estimates are small, and the 95 percent confidence interval implies we can rule out declines larger than approximately 3.5 percent of the pre-period mean.

In contrast, the overall decline in arrests is driven by reductions in less serious, non-index offenses. The largest decline is in non-index person offenses, which fall by approximately 9.4 percent relative to the pre-period mean. We also observe smaller declines in non-index property and public order offenses. These categories are generally more discretionary, consistent with a reduction in lower-level enforcement activity. Taken together, these results suggest that officers are not increasing arrests in response to the policy; instead, the observed decline reflects a pullback in lower-level, non-index enforcement rather than any shift toward more serious arrests.

4.4 Robustness

We assess robustness to alternative fixed-effects specifications and sample definitions in Figure 9. Across all outcomes, the estimated effects are highly stable in both sign and magnitude as we vary the fixed-effects structure—from saturated specifications with experiment-by-officer and experiment-by-date fixed effects to substantially more parsimonious models. This consistency indicates that our results are not driven by a particular specification or identifying variation. We also examine a subset of experiments characterized by stronger first-stage compliance, where assignment to a canceled RDO leads to substantially larger increases in

actually working. The results remain broadly similar across alternative samples and specifications, though effect sizes are modestly larger for overtime hours and for the probability of any arrest. Estimates for wellness outcomes remain centered close to zero. As expected given the smaller sample, standard errors are larger, leading to less precise estimates overall.

Second, we examine whether treatment effects might emerge with a delay, in which case our short post-treatment window could understate the true policy impact. To test this, we extend the event window to include two full work cycles before and after each cancellation, restricting to officers not treated within that horizon. As reported in Figure C.9- Figure C.11 the difference-in-differences estimates for attendance, wellness, productivity, and enforcement remain uniformly small and statistically insignificant. Together, these robustness checks confirm that our findings are not sensitive to fixed effects structure, assumptions about the persistence of treatment effects, or the length of the post-treatment window.

4.5 Treatment Heterogeneity

In the preceding sections, we described empirical evidence that the policy did not negatively impact officers across a range of measures. However, it is possible that officers with different characteristics experience the policy differently and that these heterogeneous effects are obscured when averaging over all officers. This possibility is both theoretically and empirically grounded. A large literature documents how worker productivity varies with experience, broadly, due both to selection and learning-by doing (Haggag et al. (2017), Daveri and Parisi (2015), Papay and Kraft (2015)), and police productivity specifically differs by demographic and tenure characteristics (Ba et al. (2021), Ba et al. (2022)). Under mandatory overtime, workers more susceptible to burnout have limited ability to opt out of the additional hours. For this reason, it is useful to characterize officer traits that predict larger effects.

We test for heterogeneity in the effects of mandatory overtime over a range of officer characteristics including demographics, experience, productivity, and history of canceled RDOs. We focus on these characteristics because we hypothesize that they could affect treatment compliance and susceptibility to burnout. For example, more senior officers may be better able to evade treatment, attenuating effects more than for junior officers. Officers who respond to fewer calls for service, a measure of their productivity, may be more likely to hold leadership roles and less likely to be on patrol. Then RDO cancellations may not apply to them, or if they do, they may experience less strain from working in the field in the prior week. Moreover, burnout is likely to be the result of cumulative job-related stress. RDO cancellations that occur at the start of our study period may be too early to show evidence of burnout, as officers would have experienced treatment fewer times. Alternatively, if it takes

time for officers to adjust to the policy change, it is possible we may see more evidence of a negative effect on officer behavior early on compared to when they have experienced many RDO cancellations.

In our heterogeneity analysis, we estimate a single regression that interacts the treatment with indicators for officer characteristics, allowing treatment effects to vary across subgroups while maintaining a common set of controls and fixed effects. We present the estimated subgroup-specific treatment effects in Figure 6 and Figure 7.[16] For example, we estimate the effect of RDO cancellation on officer attendance separately for male and female officers within the same regression framework.

In Figure 6, we present subgroup-specific coefficients and 95% confidence intervals for attendance and wellness outcomes: reporting to work, overtime hours, injuries, and sick days. The central pattern is the consistency of estimates across groups. For reporting to work, coefficients are negative for nearly all subgroups, but are small in magnitude and not economically meaningful. The remaining outcomes are tightly centered around zero: overtime, injuries, and sick leave estimates are modest and show limited dispersion across demographic, experience, and prior-treatment groups, with confidence intervals frequently spanning zero. Importantly, we do not observe systematic differences across subgroups that would suggest offsetting effects. Instead, the estimates move together and remain close to zero, reinforcing that the average null effects reflect genuinely small impacts rather than heterogeneity that cancels out in the aggregate.

Across calls for service, investigatory stops, arrests, and use of force (Figure 7), we again observe broadly similar patterns across subgroups. Estimates for calls for service, stops, and arrests are generally negative, though there is some dispersion and many confidence intervals include zero. We do not see evidence of meaningful positive effects for any subgroup that would offset these declines. Use-of-force estimates are consistently close to zero and imprecisely estimated across all groups. Overall, the alignment in sign and magnitude across subgroups suggests that the average effects are not driven by compositional averaging, but instead reflect modest and broadly shared changes in officer activity.

Motivated by the possibility that the effects of mandatory overtime may accumulate with repeated exposure, we next examine heterogeneity by officers' prior treatment history. While our empirical design is not optimized to isolate fully dynamic or cumulative effects, variation in prior exposure provides a useful lens for assessing whether marginal responses differ across officers with different histories of mandatory overtime. Specifically, we partition officers into

[16]Officer demographics are directly reported in rosters. Due to small sample sizes, we do not report results for officers who identify as Asian, Native/Pacific Islander, or Other. We define the categories of officer experience, tenure, and treatment history based on the median officer in our sample.

five groups based on the number of prior mandatory overtime assignments and estimate a difference-in-differences model that interacts the treatment indicator with these exposure categories. This specification allows the treatment effect to vary flexibly across officers with different histories of mandatory overtime. For each outcome, we compute group-specific marginal effects using post-estimation margins, which recover the implied treatment effect within each exposure group by combining the baseline difference-in-differences coefficient with the relevant interaction terms. We present these estimates graphically in Figure 8, along with 95% confidence intervals, to illustrate how the impact of mandatory overtime varies across the distribution of prior exposure.

Across outcomes, the estimated effects are generally small, although imprecise, with little evidence of systematic differences across exposure groups. We do not observe a clear pattern of larger effects among officers with greater prior exposure, suggesting limited evidence of cumulative or compounding effects. More broadly, these results reinforce our earlier findings that the marginal impacts of mandatory overtime are modest and relatively uniform across officers.

5 Conclusion

Overtime has become an increasingly important tool for managing chronic staffing shortages in policing and other frontline public-sector settings. In policing specifically, our results suggest that modest, strategically scheduled increases in mandatory overtime of the kind we study have little meaningful short-run effect on officer wellness, productivity, or enforcement behavior once officers return to their regular schedules. In particular, we find little evidence of subsequent changes in attendance, health-related absences, calls for service, or use of force. However, we do estimate a small decline in arrests, but the evidence indicates that this reduction is concentrated in lower-level offenses rather than more serious criminal activity.

A central implication of these findings is that the effects of overtime depend not simply on the number of additional hours worked, but also on how those hours are scheduled and what officers are asked to do during them. In our setting, the policy imposed a single additional shift that was predetermined, infrequent, and followed immediately by a day of rest. Descriptive evidence further suggests that the overtime shifts induced by the policy were generally lower intensity than ordinary patrol work and more closely resembled the types of overtime assignments officers often work voluntarily. This feature of the policy likely helps explain why the estimated effects are modest, while also clarifying the margin to which our findings most directly apply.

At the same time, our results should not be interpreted as showing that overtime is

costless or that staffing shortages can be managed indefinitely through scheduling adjustments alone. Our design identifies the short-run effects of a specific policy-induced increase in workload, not the cumulative effects of repeated overtime exposure over longer horizons. Nor do our findings speak directly to more intensive forms of mandatory overtime, to schedules that repeatedly erode opportunities for recovery, or to settings in which the additional hours are spent on more demanding forms of patrol or enforcement activity. Fatigue-related harms may still arise when overtime becomes more frequent, less predictable, or more operationally intensive.

These distinctions matter for both policy and research. For public-sector managers, our findings suggest that modest, strategically scheduled increases in overtime of this kind may be less consequential in the short run than is often assumed. More broadly, they suggest that debates over overtime should focus not only on the quantity of additional work, but also on the organizational margin through which that work is deployed. The consequences of overtime are likely to differ depending on whether agencies use it to staff ordinary patrol, extend active enforcement, or cover lower-intensity visibility-oriented assignments. For researchers, our results underscore the importance of distinguishing between discrete scheduling shocks and cumulative workload exposure when evaluating the effects of labor scarcity on worker performance and wellbeing. Understanding when additional hours begin to generate meaningful deterioration in police productivity, health, morale, or enforcement outcomes remains an important question for future work.

References

Ba, B., P. Bayer, N. Rim, R. Rivera, and M. Sidibé (2022). Police officer assignment and neighborhood crime. *NBER Working Paper 29243*.

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.

Baltimore Police Department (2023). Bpd response to office of legislative audits overtime report.

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.

Chalfin, A. and F. Goncalves (2020). The pro-social motivations of police officers. Technical report, Working Paper, National Bureau of Economic Research November.

Chicago Police Department (2023). General order g03-02-02: Incidents requiring the completion of a tactical response report. `https://directives.chicagopolice.org/#directive/public/6610`. Accessed [August 12, 2025].

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

City of Chicago Office of Inspector General (2022, August). Consecutive days worked by cpd members, april-may 2022.

Daveri, F. and M. L. Parisi (2015). Experience, innovation, and productivity: Empirical evidence from italy's slowdown. *ILR Review 68*(4), 889–915.

de Chaisemartin, C. and X. D'Haultfoeuille (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal 26*, C1–C30.

Deshpande, M. and Y. Li (2019). Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy 11*(4), 213–248.

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.

Fan, W., J. B. Schor, O. Kelly, and G. Gu (2025). Work time reduction via a 4-day workweek improves worker well-being. *Nature Human Behaviour*. Forthcoming, published online July 2025.

Ferrazares, T. (2025). Shift structure and cognitive depletion: Evidence from police officers. Working Paper. UC Santa Barbara. `https://ferrazares.github.io/assets/FERRAZARESjmp.pdf`.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics 225*(2), 254–277. Themed Issue: Treatment Effect 1.

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.

Haggag, K., B. McManus, and G. Paci (2017, January). Learning by driving: Productivity improvements by new york city taxi drivers. *American Economic Journal: Applied Economics 9*(1), 70–95.

Holz, J. E., R. G. Rivera, and B. A. Ba (2023). Peer effects in police use of force. *American Economic Journal: Economic Policy 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. *Police Quarterly 21*(1), 30–52.

Kc, D. S. and C. Terwiesch (2009). Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *55*(9), 1486–1498.

Kuntz, L., R. Mennicken, and S. Scholtes (2015). Stress on the ward: Evidence of safety tipping points in hospitals. *61*(4), 754–771.

Linos, E., K. Ruffini, and S. Wilcoxen (2022). Reducing burnout and resignations among frontline workers: A field experiment. *Journal of Public Administration Research and Theory 32*(3), 473–488.

Lu, S. F. and L. X. Lu (2017, November). Do mandatory overtime laws improve quality? staffing decisions and operational flexibility of nursing homes. *Management Science 63*(11), 3566–3585.

Maryland Office of Legislative Audits (2023). Performance audit of overtime policies and practices at the baltimore police department. Covers December 2020–June 2022.

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.

Masterson, M. (2024, June). Chicago spent more than 1 million overtime hours on 'scarecrow' police shifts in 2022 before abandoning strategy. Accessed: 2025-12-03.

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.

NBC Chicago (2023, October). New efforts to boost officer morale in cpd gives time back to officers. Accessed: 2025-12-03.

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.

New York City Department of Investigation (2023a). An analysis of nypd overtime and risks to officer and public safety. Report, accessed July 2024.

New York City Department of Investigation (2023b). An assessment of nypd overtime, fatigue, and off-duty employment.

Owens, E. and B. A. Ba (2021). The economics of policing and public safety. *Journal of Economic Perspectives 35*(4), 3–28.

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.

Papay, J. P. and M. A. Kraft (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics 130*, 105–119.

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 Francisco Chronicle (2024a). Here's how sfpd wants to cut down on overtime. *San Francisco Chronicle*.

San Francisco Chronicle (2024b). Sfpd to get massive overtime boost despite concerns about spending. *San Francisco Chronicle*.

Schuba, T. and F. Main (2024, January). Cpd supt. snelling: 'scarecrow' policing in loop, downtown is over. Accessed: 2025-12-03.

Sisk, C. (2023, October). New efforts to boost officer morale in cpd gives time back to officers. CBS Chicago, accessed September 11, 2024.

Wing, C., S. M. Freedman, and A. Hollingsworth (2023). Stacked difference-in-differences. *National Bureau of Economic Research Working Paper Series* (32054).

6 Tables

Table 1: Officer Activity Across Patrol and Overtime Assignments

	Any			Count			
	Arrests	Stops	Calls	Arrests	Stops	Calls	N (Officer-Days)
Panel A: Any Overtime							
Voluntary RDO (>0 hrs)	0.036	0.080	0.377	0.095	0.059	2.332	60,301
Involuntary RDO (>0 hrs)	0.022	0.051	0.186	0.065	0.037	0.873	36,820
Extended OT (>0 hrs)	0.148	0.298	0.738	0.415	0.262	5.547	114,437
Normal Hours (no OT)	0.057	0.114	0.714	0.136	0.091	4.143	698,963
Panel B: Long Overtime (>4 hrs)							
Voluntary RDO (>4 hrs)	0.036	0.080	0.377	0.095	0.059	2.324	59,884
Involuntary RDO (>4 hrs)	0.022	0.051	0.186	0.065	0.037	0.873	36,714
Extended OT (>4 hrs)	0.169	0.393	0.644	0.590	0.284	4.652	21,168

This table reports mean outcomes at the officer-day level by overtime status. Voluntary RDO and involuntary RDO refer to overtime worked on scheduled days off. Extended OT refers to overtime worked beyond regular shifts. Normal Hours includes officer-days with positive work activity and no overtime. Outcomes include indicators for any stop, any arrest, and any call for service (CFS), as well as total counts.

Table 2: Summary Statistics

	Linked Sample		Stacked Sample	
VARIABLES	Mean	SD	Mean	SD
Reports to Work	.5382	.4985	.5455	.4979
Total Overtime Hours	.9277	2.551	.7835	2.305
Total Involuntary Overtime Hours	.1952	1.298	.1636	1.198
Total Calls for Service	1.88	3.512	2.373	3.856
Total Investigatory Stops	.04466	.316	.05971	.3663
Any Arrests	.0587	.2351	.07468	.2629
Any Use of Force	.0012	.0356	.001749	.04287
Female	.2037	.4028	.2202	.4144
White	.6845	.4647	.6838	.465
Hispanic	.346	.4757	.3872	.4871
Black	.1477	.3548	.1649	.3711
Years of Experience	12.29	8.869	9.843	8.053
N	3,152,660		1,854,336	
Injured on the Job	.03508	.184	.04412	.2054
Calls in Sick	.05634	.2306	.06146	.2402
N	3,147,027		1,850,441	
Distinct Officers	7,927	-	5,864	-
Distinct Sub-Experiments	-	-	288	-
Officers per Sub-Experiment	-	-	623.4	146.8

Table 3: Effects of Mandatory Overtime on Officer Activity, Attendance, and Wellness

	(A) Attendance and Work		(B) Officer Wellness	
Dependent Variable:	Reports to Work	Overtime Hours	On the Job Injured	Calls in Sick
Treatment × Post	-0.00658***	-0.00836	0.00017	0.00029
	(0.00131)	(0.00763)	(0.00029)	(0.00063)
Pre-Period Mean	0.6559	0.5513	0.0419	0.0641
Percent of Mean	-1.00	-1.52	0.41	0.45
Observations	1406664	1406664	1406664	1406664
95% CI (% of Mean)	[-1.40, -0.61]	[-4.23, 1.20]	[-0.95, 1.76]	[-1.48, 2.38]
	(C) Officer Productivity		**(D) Enforcement Activity**	
Dependent Variable:	Calls for Service	Investigatory Stops	Any Arrests	Any Use of Force
Treatment × Post	-0.02498	-0.00262	-0.00220*	-0.00004
	(0.01573)	(0.00171)	(0.00129)	(0.00018)
Pre-Period Mean	2.8652	0.0740	0.0906	0.0020
Percent of Mean	-0.87	-3.54	-2.43	-1.87
Observations	1406664	1406664	1406664	1406664
95% CI (% of Mean)	[-1.95, 0.20]	[-8.07, 0.99]	[-5.22, 0.36]	[-19.65, 15.65]

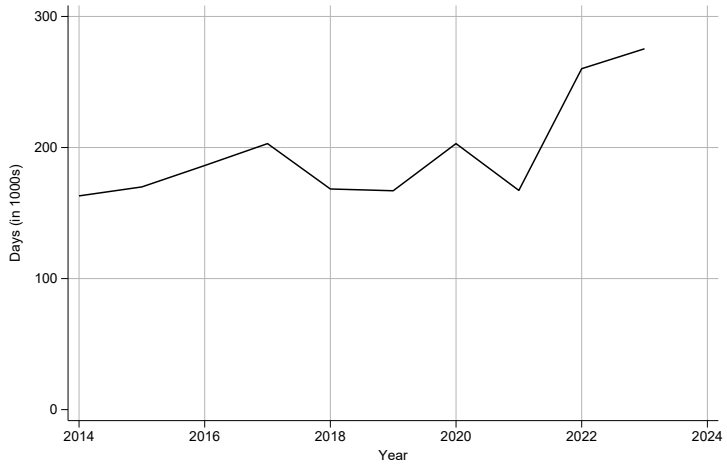
Data are restricted to exclude observations falling on RDOs, specifically with event-times $t \in \{-5, 0, 1\}$. Results are estimated using an aggregated version of our dynamic event-study specification, Equation 1. We report the coefficients on interaction between treatment and a post-treatment indicator, omitting coefficients for event-times $t < 1$. Standard errors clustered by area-DOG are reported in parentheses. The pre-treatment mean of each outcome is calculated for event-times $t < 0$, excluding RDOs.

Table 4: Effects of Mandatory Overtime on Arrests by Offense Type

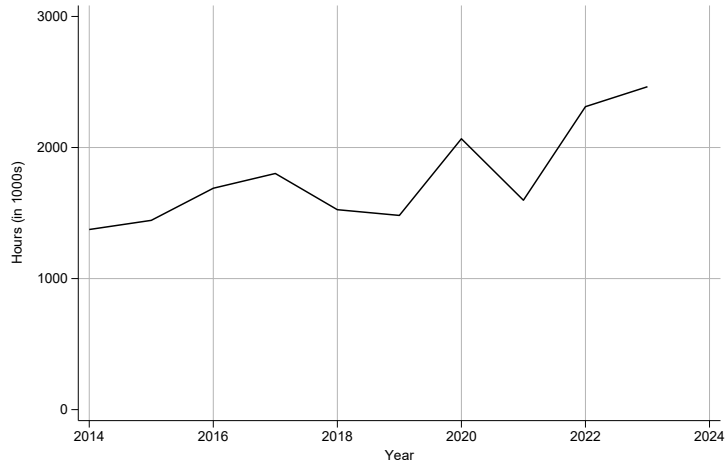
Dependent Variable:	Index Arrests	Traffic Arrests	Non-Index Property	Non-Index Personal	Non-Index Public Order	No-Charge Arrests
Treatment × Post	0.00077	-0.00015	-0.00065	-0.00178**	-0.00125	0.00040
	(0.00073)	(0.00091)	(0.00062)	(0.00085)	(0.00138)	(0.00044)
Pre-Period Mean	0.0186	0.0265	0.0187	0.0189	0.0533	0.0036
Percent of Mean	4.16	-0.57	-3.49	-9.42	-2.35	11.07
Observations	1406664	1406664	1406664	1406664	1406664	1406664
95% CI	[-3.55, 11.83]	[-7.30, 6.16]	[-9.97, 3.02]	[-18.23, -0.60]	[-7.42, 2.73]	[-12.84, 35.07]

Data are restricted to exclude observations falling on RDOs, specifically with event-times $t \in \{-5, 0, 1\}$. Results are estimated using an aggregated version of our dynamic event-study specification, Equation 1. We report the coefficients on interaction between treatment and a post-treatment indicator, omitting coefficients for event-times $t < 1$. Standard errors clustered by area-DOG are reported in parentheses. The pre-treatment mean of each outcome is calculated for event-times $t < 0$, excluding RDOs.

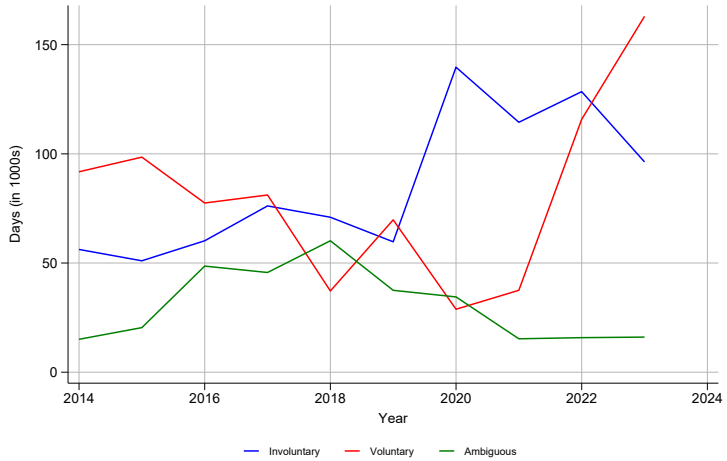
7 Figures



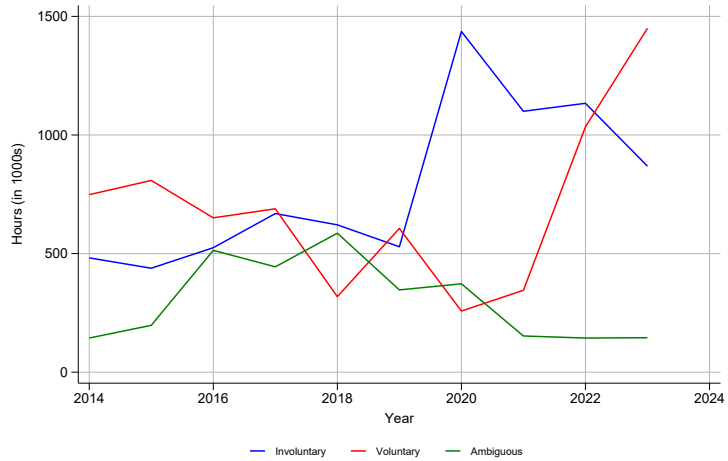
(a) Cumulative Days with Overtime



(b) Cumulative Hours of Overtime



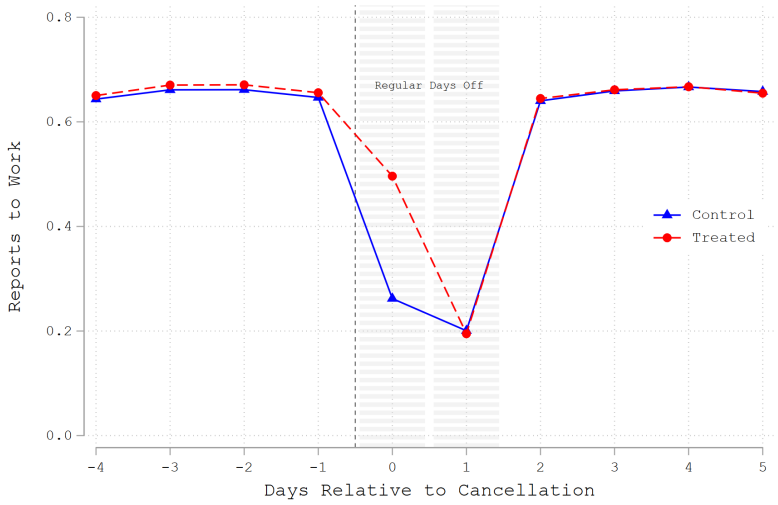
(c) Cumulative Days with Overtime by Type



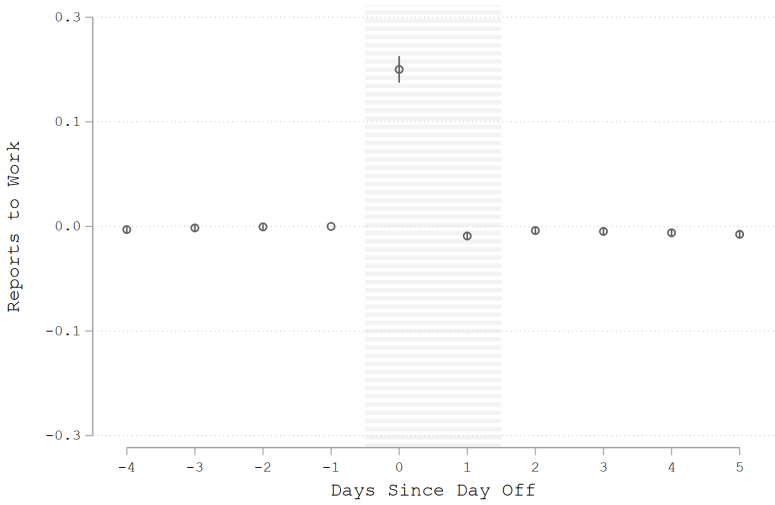
(d) Cumulative Hours of Overtime by Type

Figure 1: Trends in Overtime from 2014 to 2023

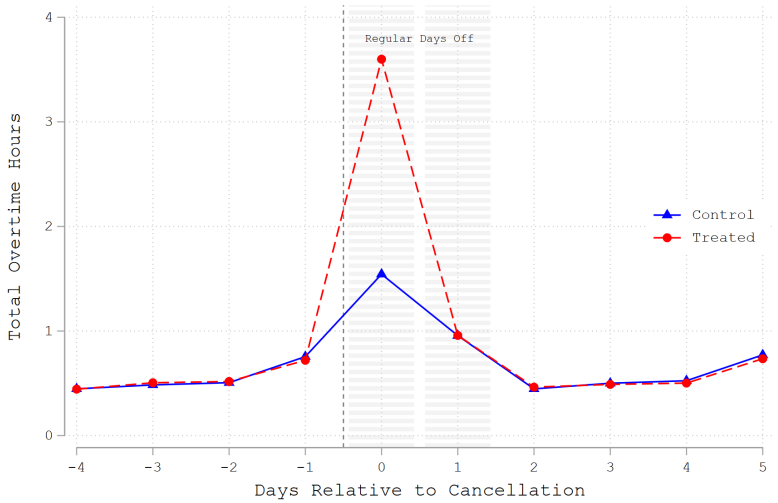
These figures show trends in overtime usage per officer. The data is restricted to overtime assignments where an officer worked 8 or more hours. Panel (a) plots cumulative days with an officer taking any overtime since 2014, panel (b) plots cumulative hours of overtime since 2014, panel (c) plots cumulative days of overtime by type since 2014, and panel (d) plots cumulative hours of overtime by type since 2014.



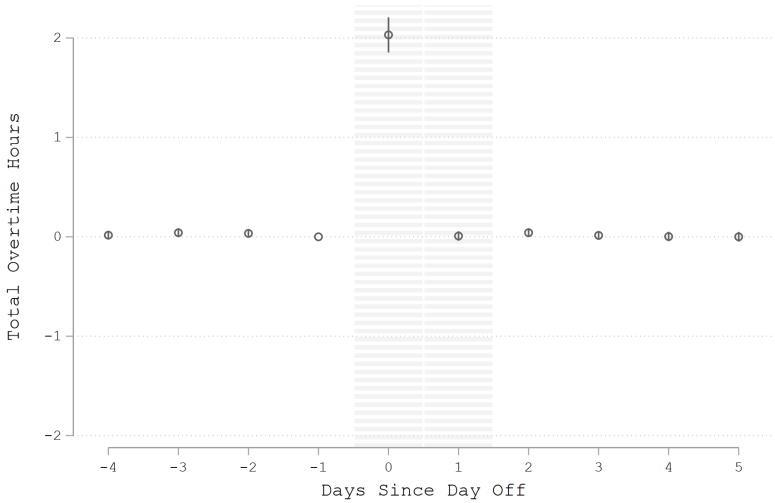
(a) Raw Data: Reported to Work



(b) Event Study: Reported to Work



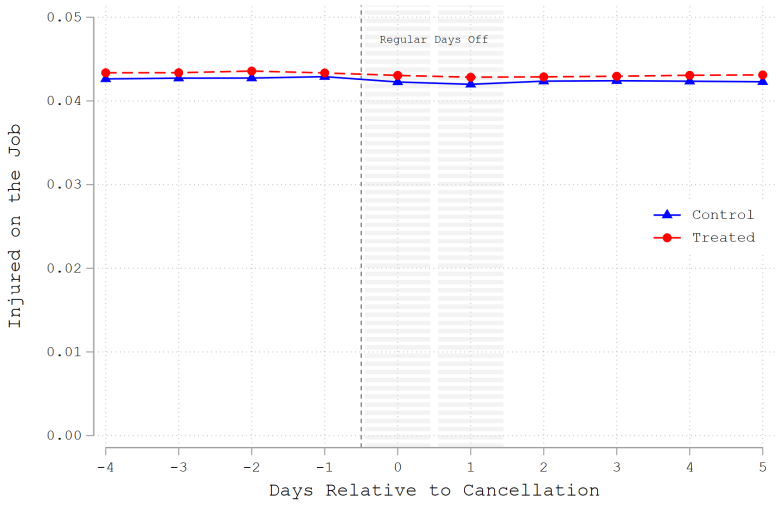
(c) Raw Data: Overtime Hours



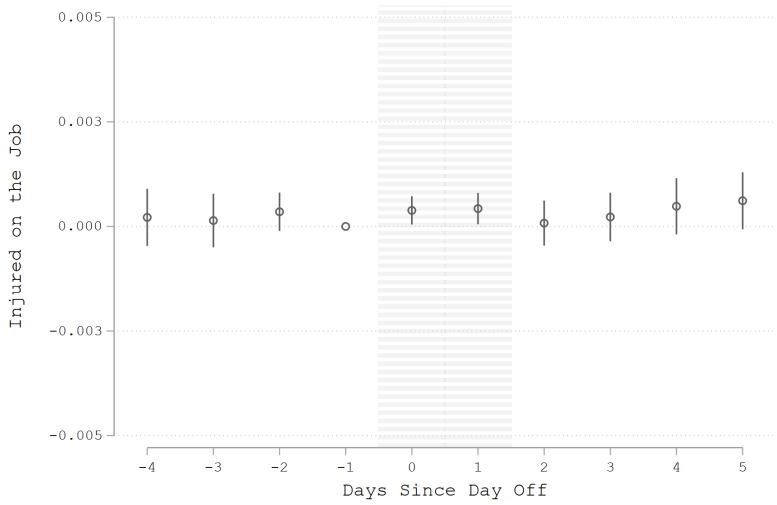
(d) Event Study: Overtime Hours

Figure 2: Mandatory Overtime: Effects on Officer Attendance and Hours Worked

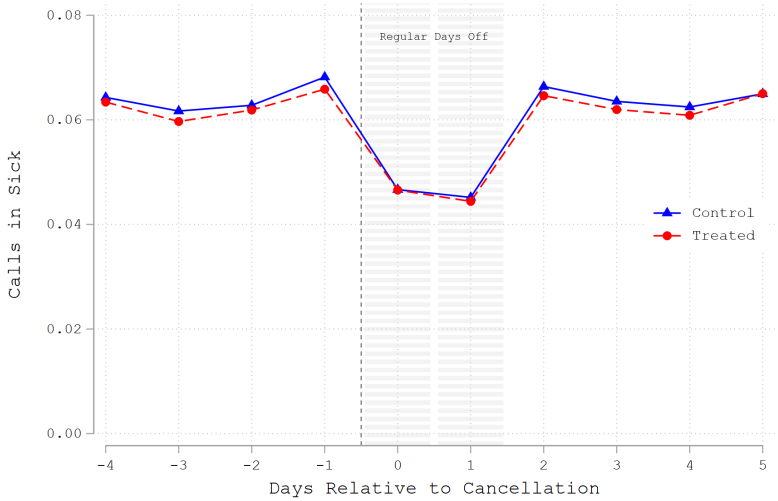
These figures show the effect of overtime assignments on daily officer attendance and overtime hours. Panels (a) and (c) present raw binscatter evidence, while panels (b) and (d) display event study estimates from Equation 1.



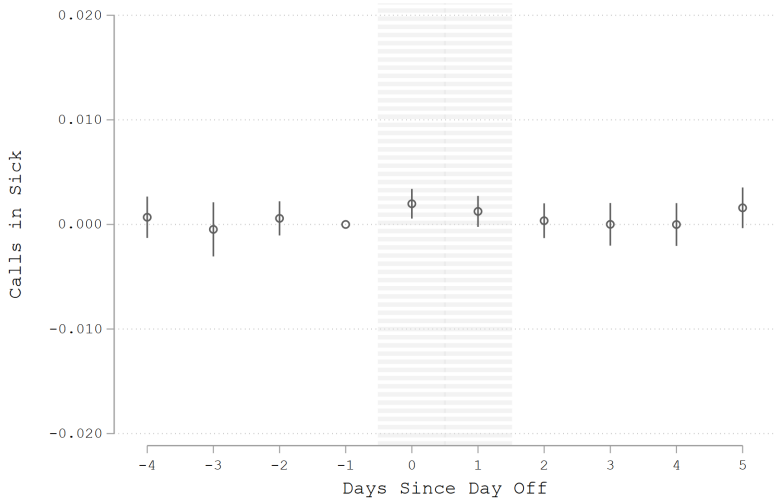
(a) Raw Data: Injury



(b) Event Study: Injury

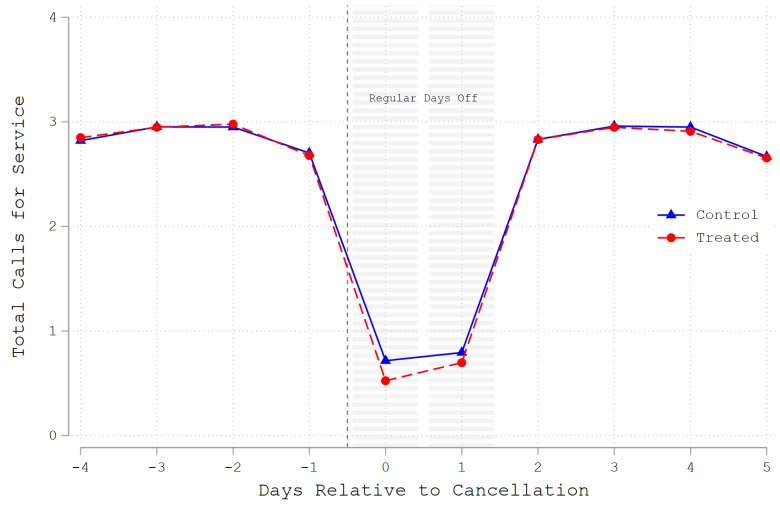


(c) Raw Data: Calls in Sick

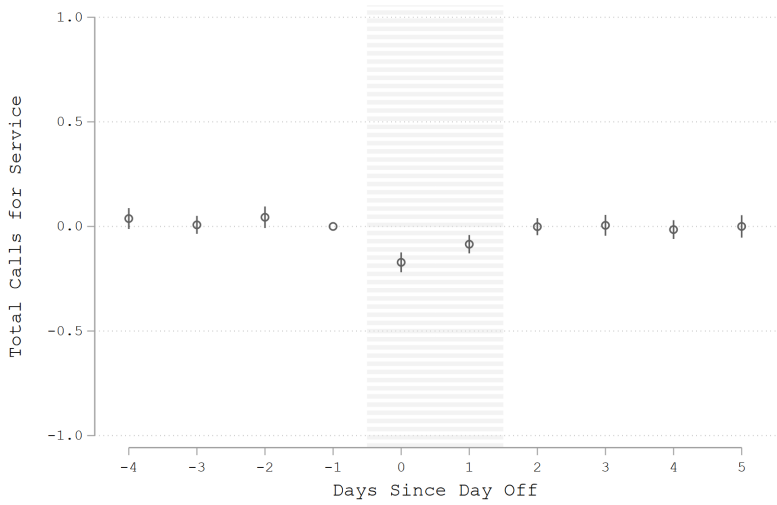


(d) Event Study: Calls in Sick

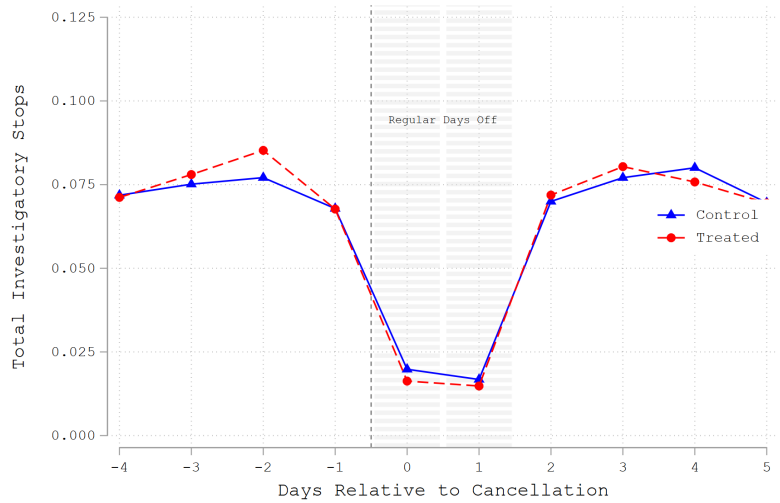
Figure 3: Mandatory Overtime: Effects on Officer Injuries and Sick Leave



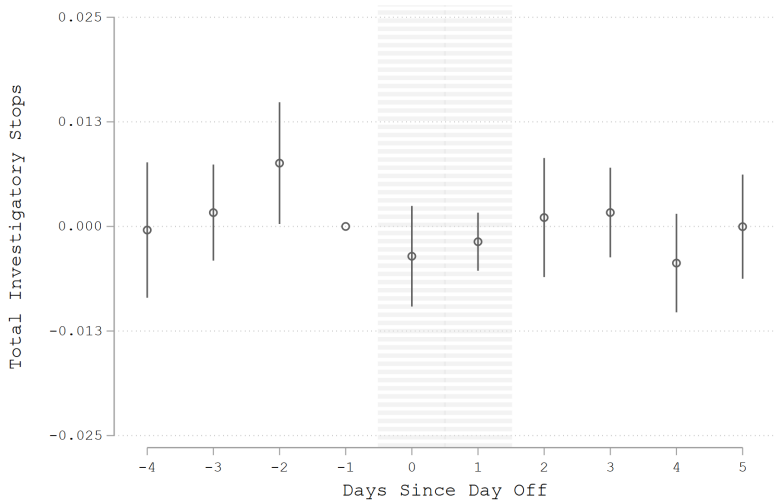
(a) Raw Data: Calls for Service



(b) Event Study: Calls for Service

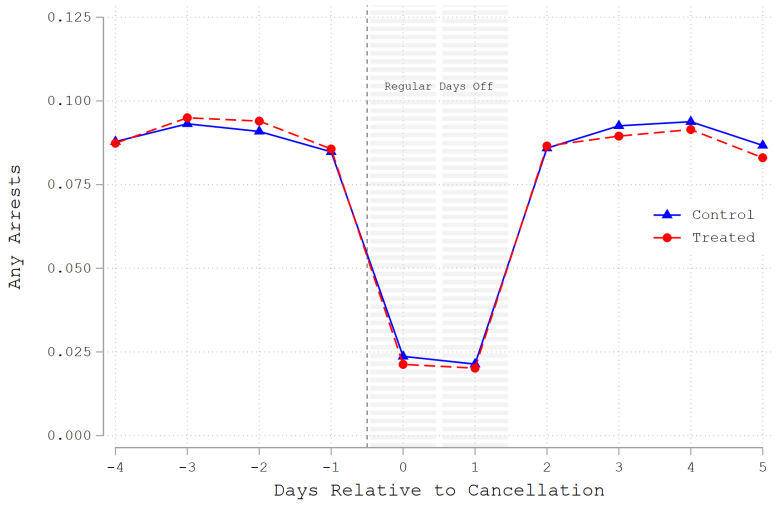


(c) Raw Data: Stops

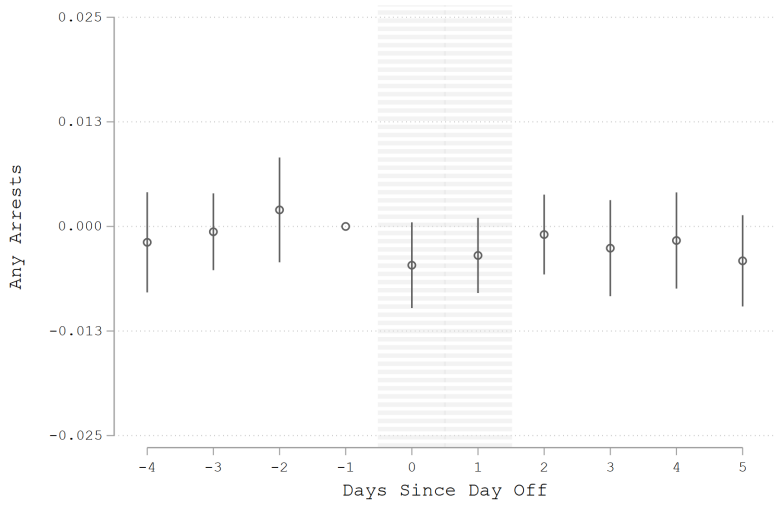


(d) Event Study: Stops

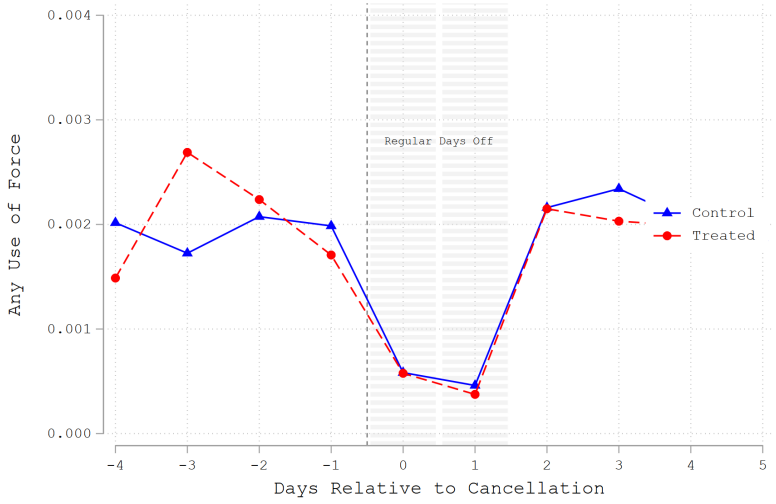
Figure 4: Mandatory Overtime: Effects on Calls for Service and Stops



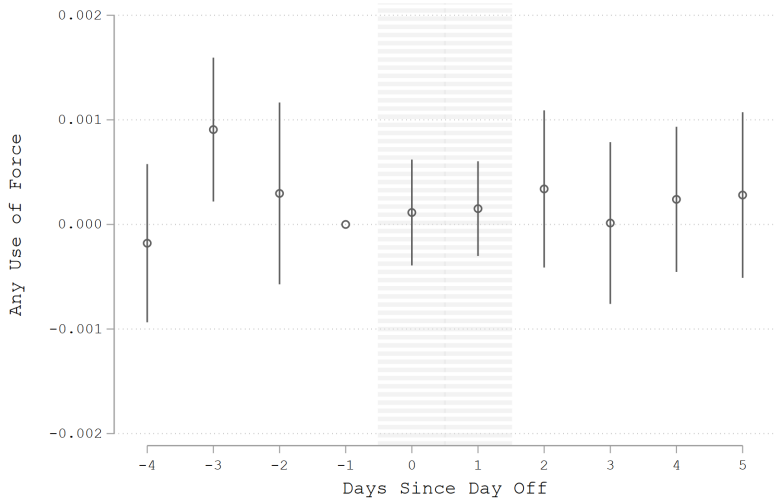
(a) Raw Data: Arrests



(b) Event Study: Arrests



(c) Raw Data: Use of Force



(d) Event Study: Use of Force

Figure 5: Mandatory Overtime: Effects on Arrests and Use of Force

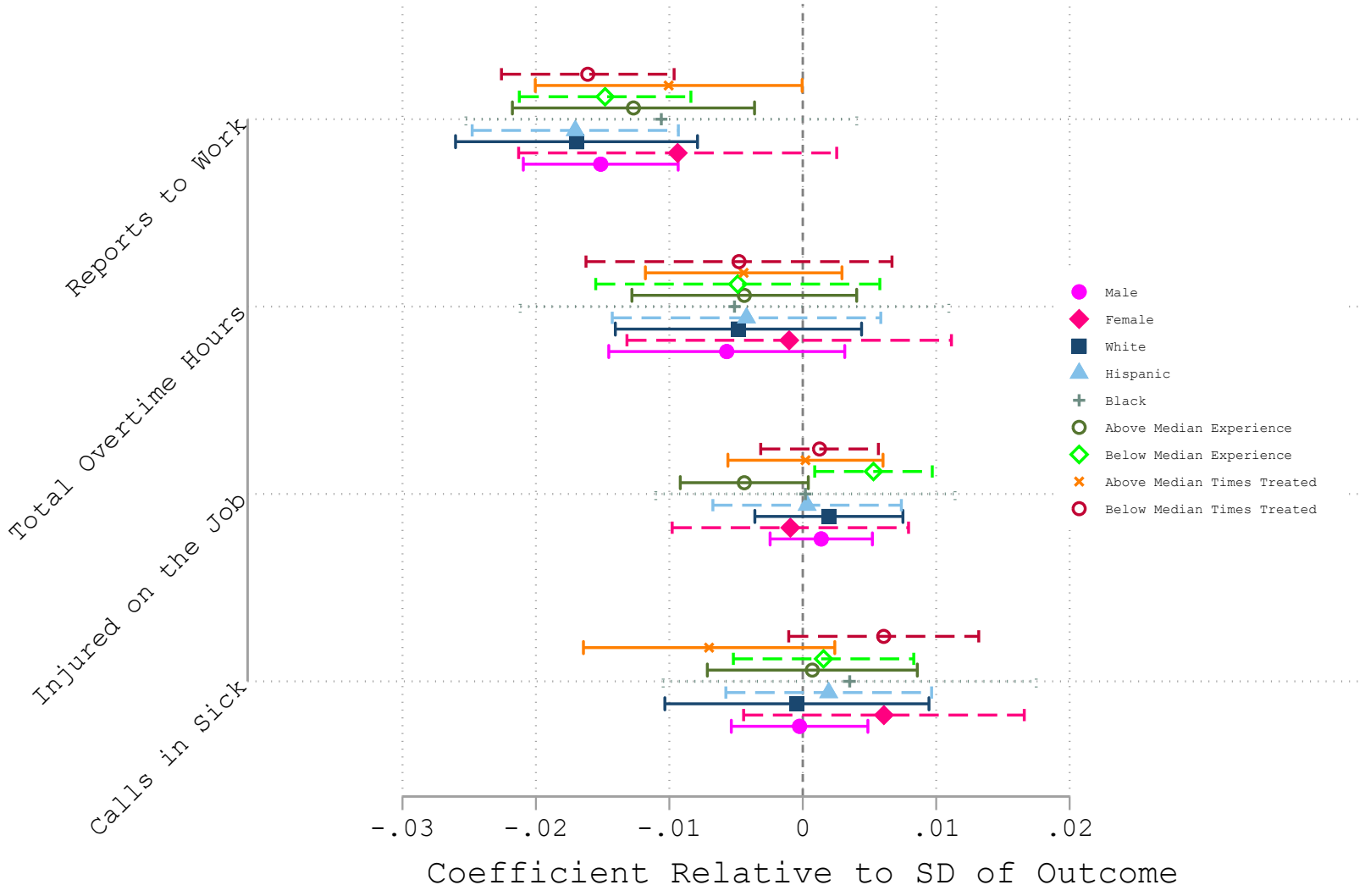


Figure 6: Heterogeneous Effects of Mandatory Overtime on Attendance and Wellness Outcomes

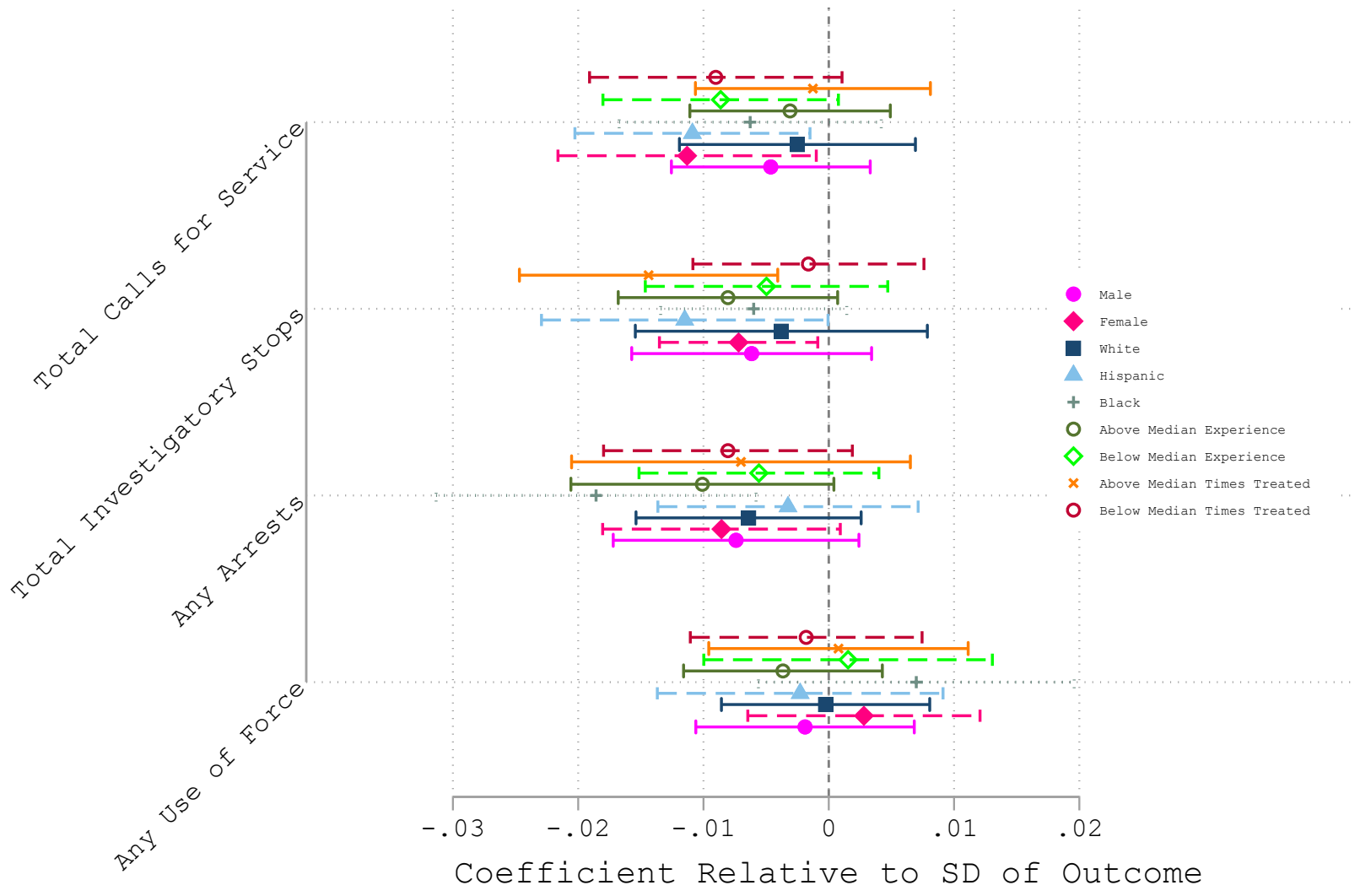
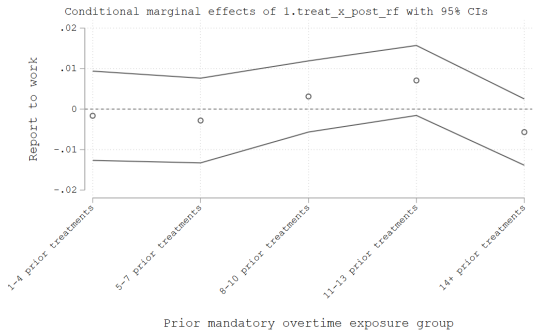
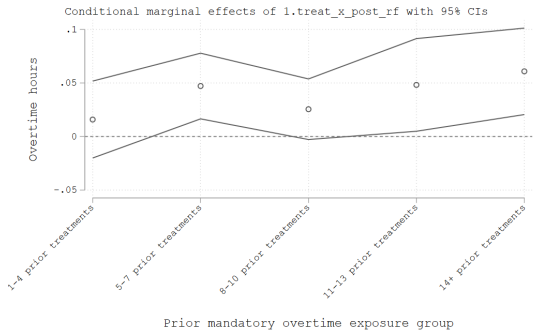


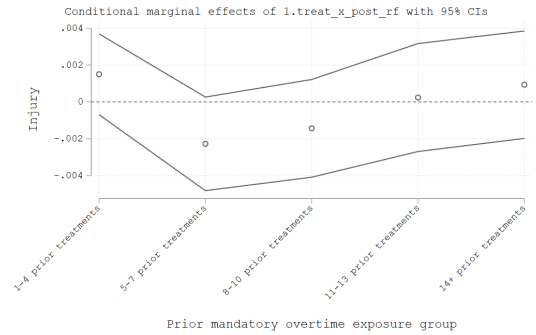
Figure 7: Heterogeneous Effects of Mandatory Overtime on Enforcement and Productivity Outcomes



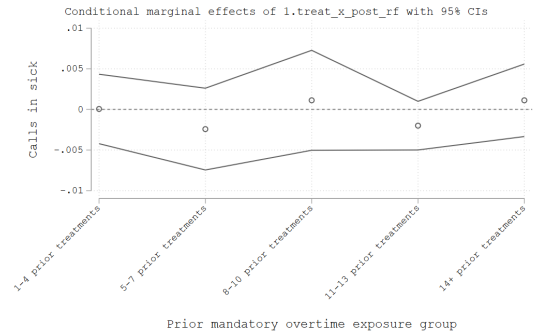
(a) Report to work



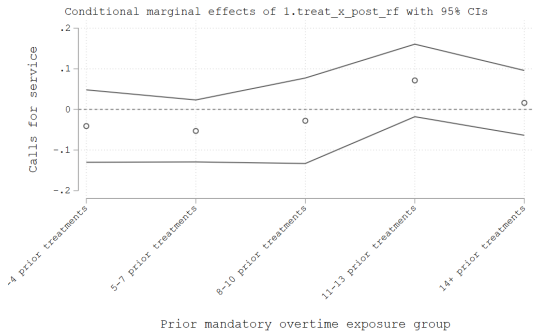
(b) Overtime hours



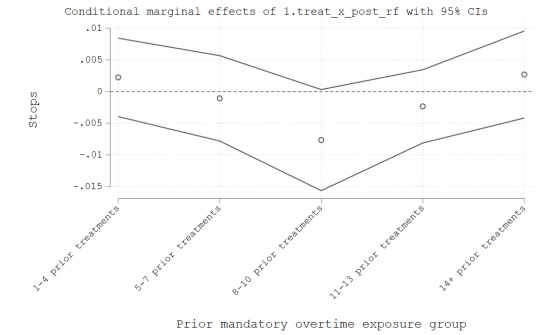
(c) Injury



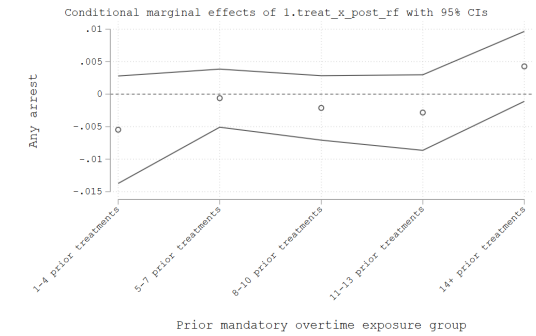
(d) Calls in sick



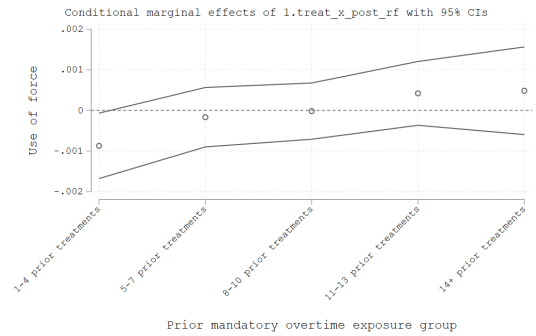
(e) Calls for service



(f) Stops



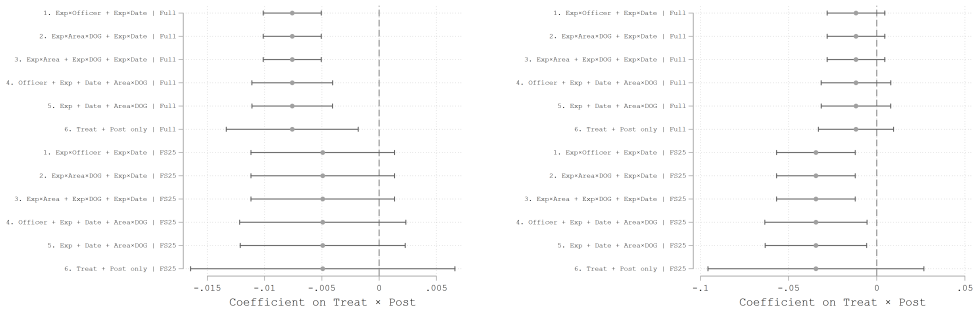
(g) Any arrest



(h) Use of force

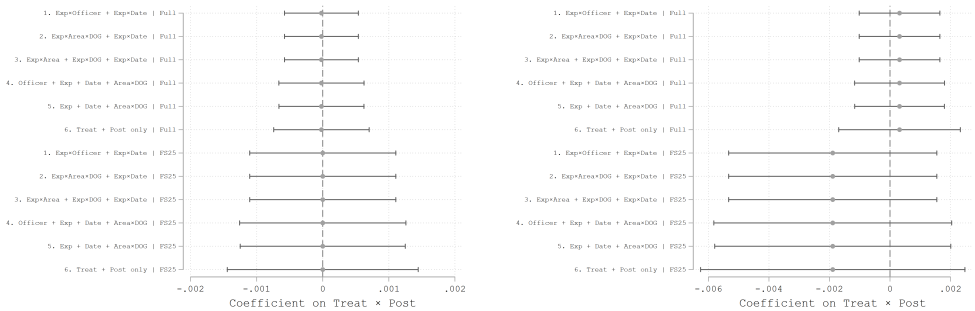
Figure 8: Heterogeneous Effects of Mandatory Overtime by Prior Exposure to Overtime

Points show difference-in-differences estimates within each exposure group; bars denote 95% confidence intervals.



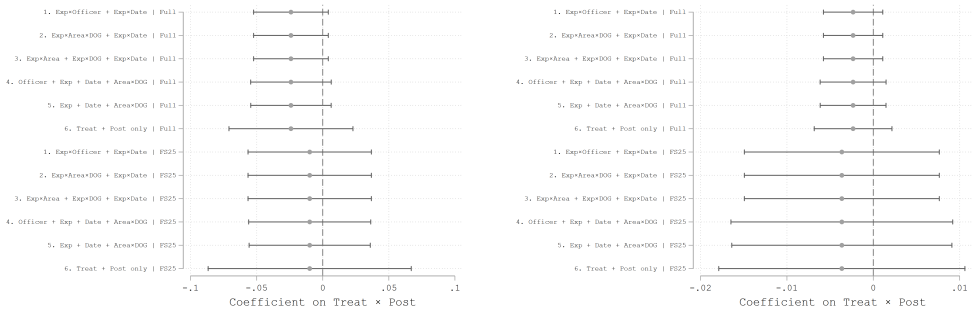
(a) Attendance

(b) Overtime Hours



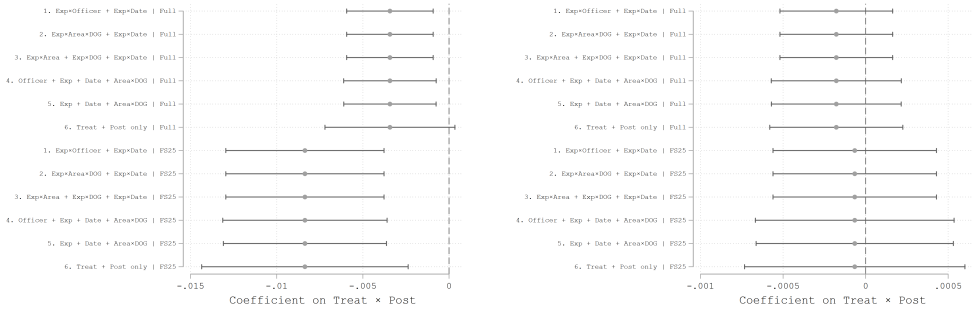
(c) Injuries

(d) Sick Leave



(e) Calls for Service

(f) Investigatory Stops



(g) Any Arrest

(h) Use of Force

Figure 9: Robustness of Treatment Effects Across Fixed Effects Specifications and Samples

Estimated effects of mandatory overtime across alternative fixed-effects specifications and samples. Each panel plots the coefficient on Treat × Post and 95% confidence intervals from regressions that vary the included fixed effects. The top six estimates use the full sample, while the bottom six restrict to the FS25 sample, which includes experiments with a first-stage effect above the 25th percentile.

A Data Appendix

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 they adhered to the stable six day work cycle described in the main text, and were subject to the RDO cancelation policy of interest. We also restrict our sample to officers assigned areas numbered 1-5. While The CPD does assigns some officers to alternative "areas", 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 A.10 shows an example of an RDO cancellation matrix we obtained from the CPD through a FOIA request. RDO cancelations have a sequential, cyclic structure, with occasional exceptions. For example, on May 31, 2022, the cancellation cycle begins with officers in Area 1 and DOG 61[17]. Comparing this figure to the operations calendar for 2022 confirms 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.

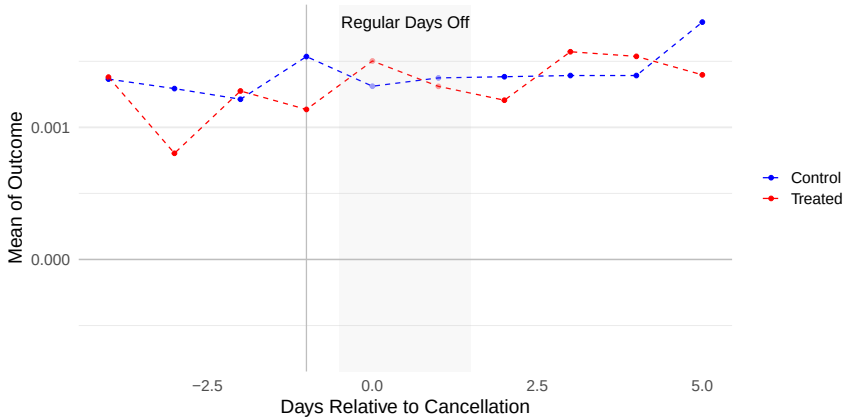
[17]DOG 75 also experiences an RDO cancellation on this date, but our analysis is restricted to officers in DOGs 61 to 66. See 2.2 for details on this sample restriction.

Dates	RDO	District Law Enforcement				
		Area 1	Area 2	Area 3	Area 4	Area 5
Tues, 31-May	61, 75	X				
Wed, 01-Jun	62, 76	X				
Thurs, 02-Jun	63, 77	X				
Fri, 3-Jun	64, 71, F/S	X				
Sat, 4-Jun	65, 72, S/S	X				
Sun, 5-Jun	66, 73, S/M	X				
Tues, 07-Jun	62, 74			X	X	
Wed, 08-Jun	63, 75			X	X	
Thurs, 09-Jun	64, 76			X	X	
Fri, 10-Jun	65, 77, F/S			X	X	
Sat, 11-Jun	66, 71, S/S			X	X	
Sun, 12-Jun	61, 72, S/M			X	X	
Tues,21-Jun	64, 72		X			X
Wed, 22-Jun	65, 73		X			X
Thurs, 23-Jun	66, 74		X			X
Fri, 24-Jun	61, 75, F/S		X			X
Sat, 25-Jun	62, 76, S/S		X			X
Sun, 26-Jun	63, 77, S/M		X			X
Tues, 05-Jul	66, 77	X				
Wed, 06-Jul	61, 71	X				
Thurs, 07-Jul	62, 72	X				
Fri, 8-Jul	63, 73, F/S	X				
Sat, 9-Jul	64, 74, S/S	X				
Sun, 10-Jul	65,75, S/M	X				
Tues, 12-Jul	61, 76			X	X	
Wed, 13-Jul	62, 77			X	X	
Thurs, 14-Jul	63, 71			X	X	
Fri, 15-Jul	64, 72, F/S			X	X	
Sat, 16-Jul	65, 73, S/S			X	X	
Sun, 17-Jul	66, 74, S/M			X	X	
Tues, 19-Jul	62, 75		X			X
Wed, 20-Jul	63, 76		X			X
Thurs, 21-Jul	64, 77		X			X
Fri, 22-Jul	65, 71, F/S		X			X
Sat, 23-Jul	66, 72, S/S		X			X
Sun, 24-Jul	61, 73, S/M		X			X

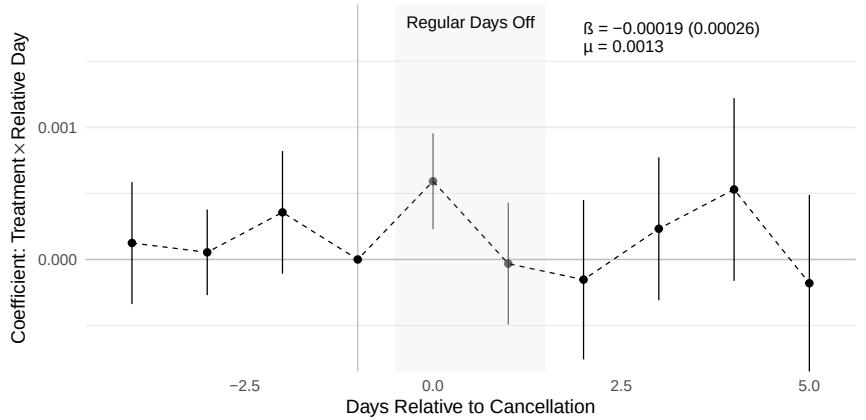
Figure A.10: Example CPD Cancellation Matrix

B Appendix Figures

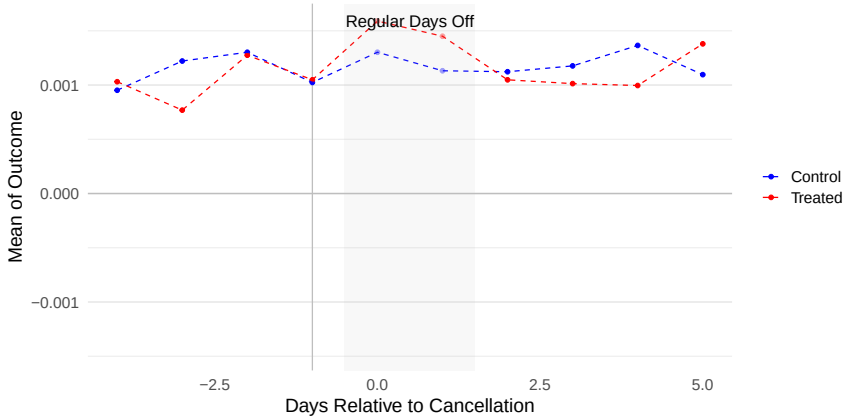
Figure C.1: *Other Outcomes* Work Assignment



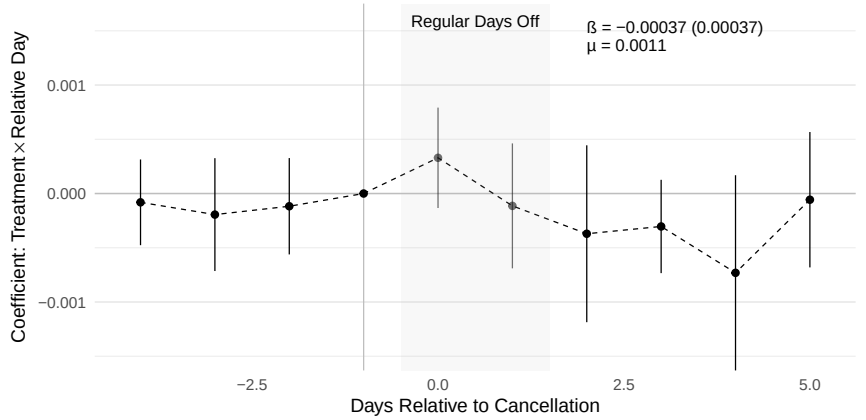
(a) Raw Data: DOG Assignment



(b) Event Study: DOG Assignment

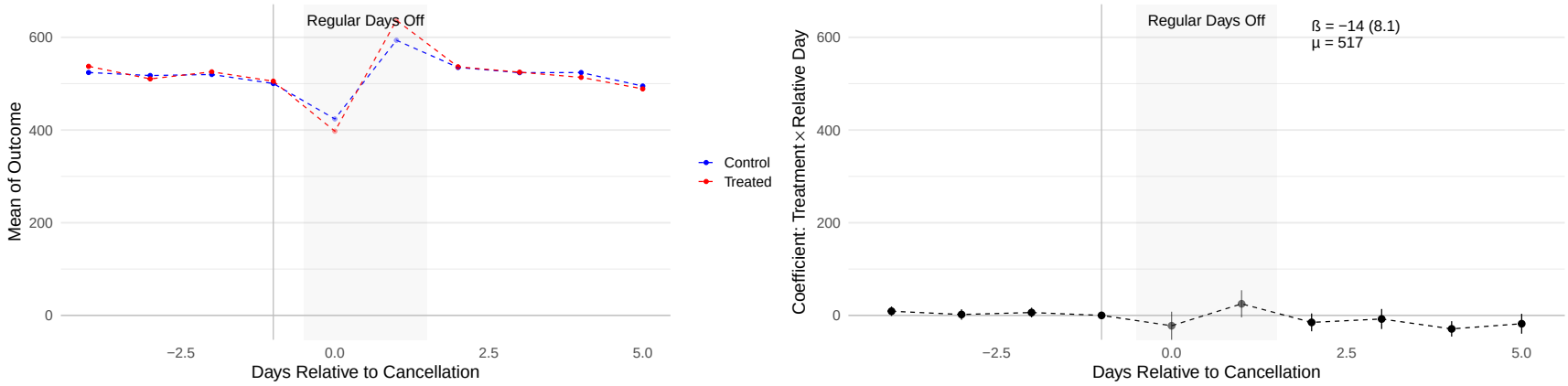


(c) Raw Data: Area Assignment



(d) Event Study: Area Assignment

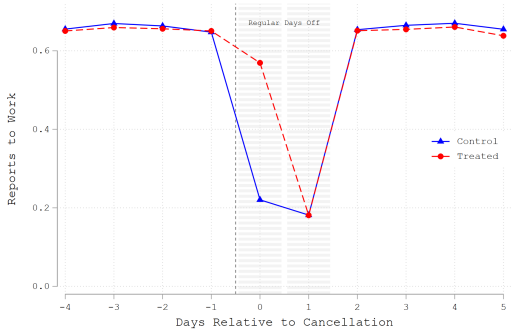
Figure C.2: *Other Outcomes* Time from Call Received by Dispatcher to On-Scene in Seconds



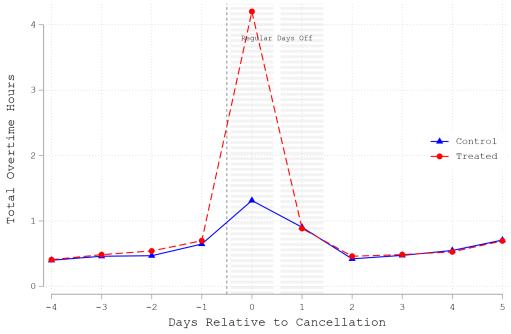
(a) Binscatter Plot

(b) Dynamic DID Event Study

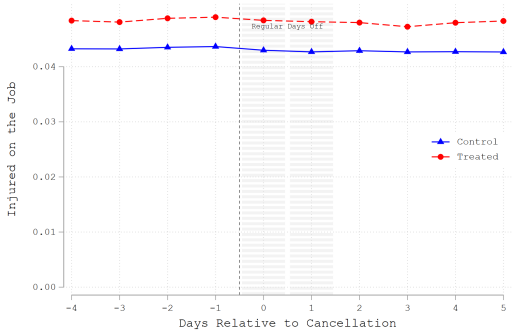
Figure C.3: *Robustness* Strong First Stage Sample



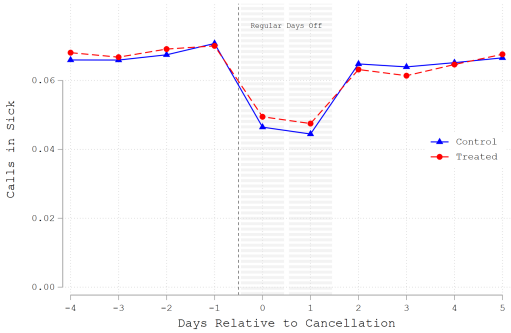
(a) Reports to Work



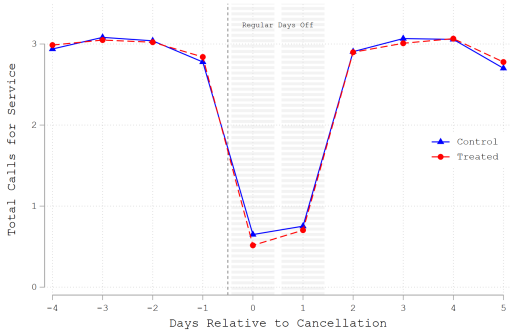
(b) Overtime Hours



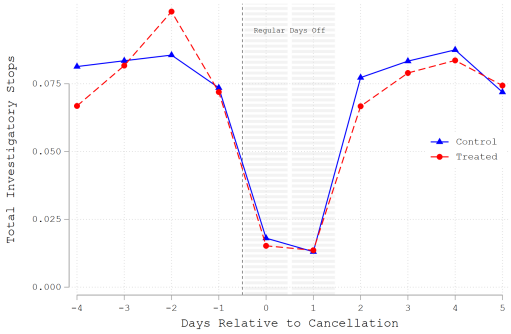
(c) Injured on the Job



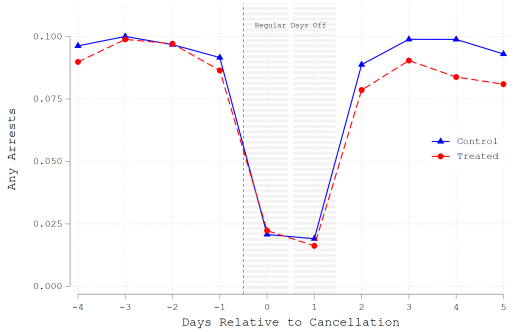
(d) Calls in Sick



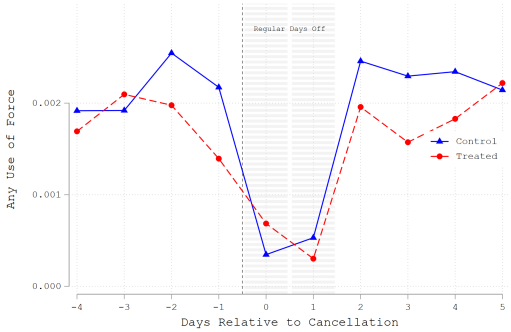
(e) Calls for Service



(f) Investigatory Stops



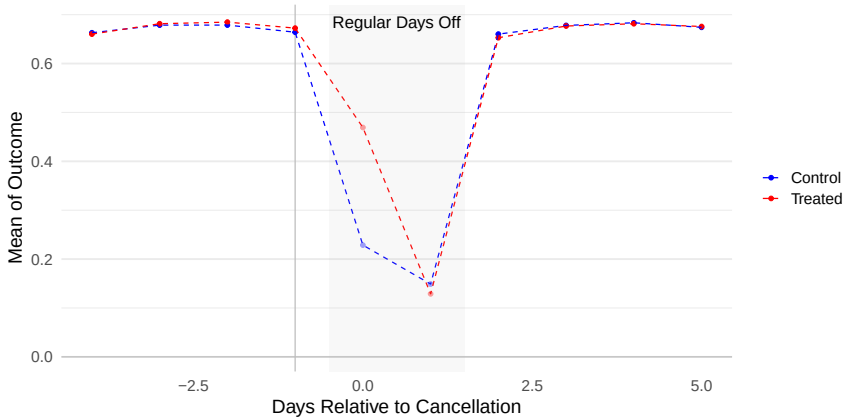
(g) Any Arrest



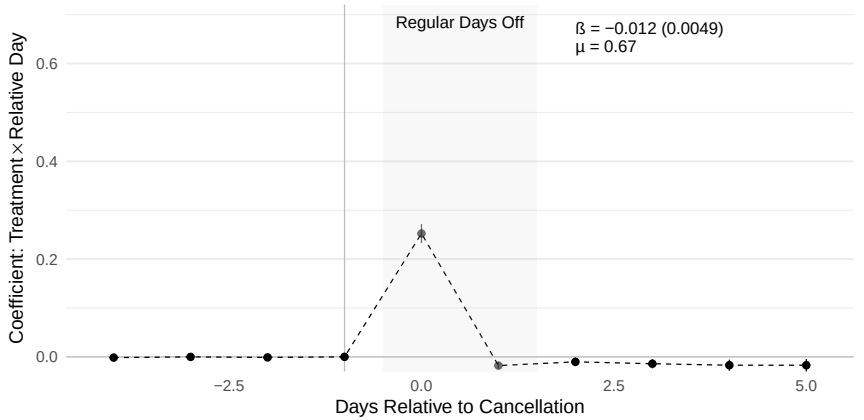
(h) Any Use of Force

Raw means for treated and control officers in experiments with the strongest first stages, plotted relative to the RDO cancellation date.

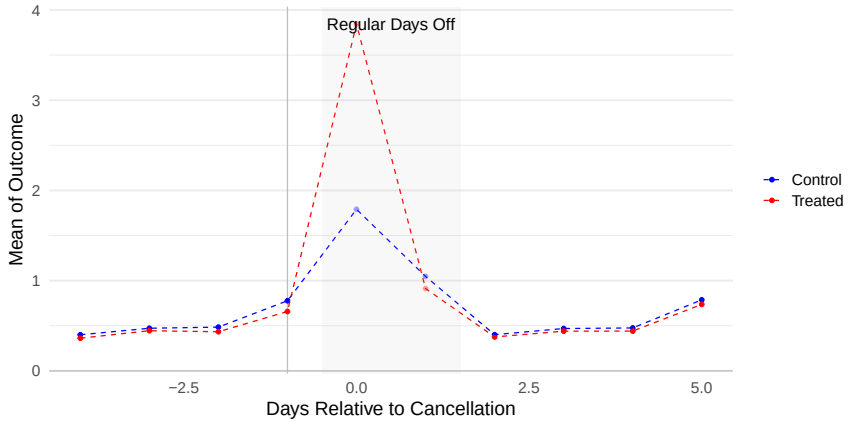
Figure C.4: *Robustness* Restricted Sample and Attendance



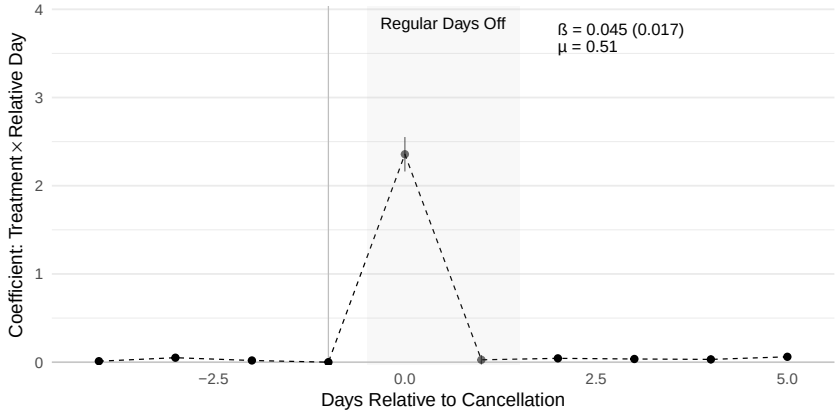
(a) Raw Data: At Work



(b) Event Study: At Work

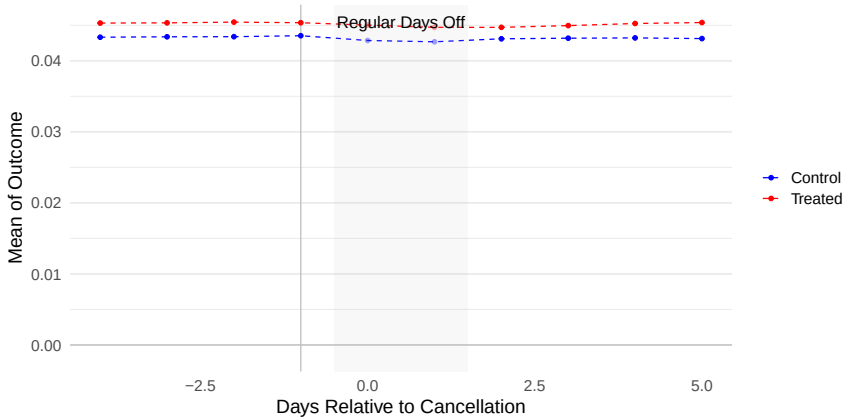


(c) Raw Data: Overtime Hours

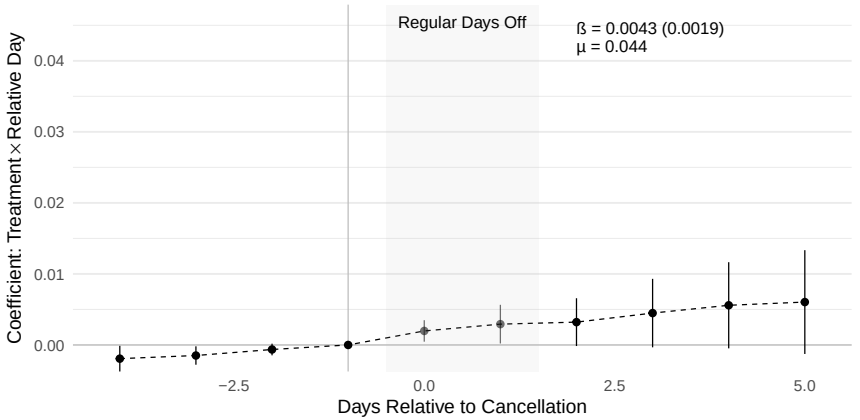


(d) Event Study: Overtime Hours

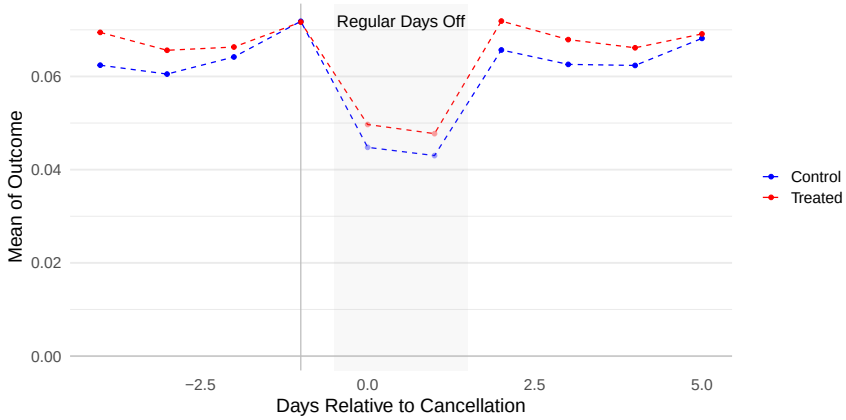
Figure C.5: *Robustness* Restricted Sample and Wellness



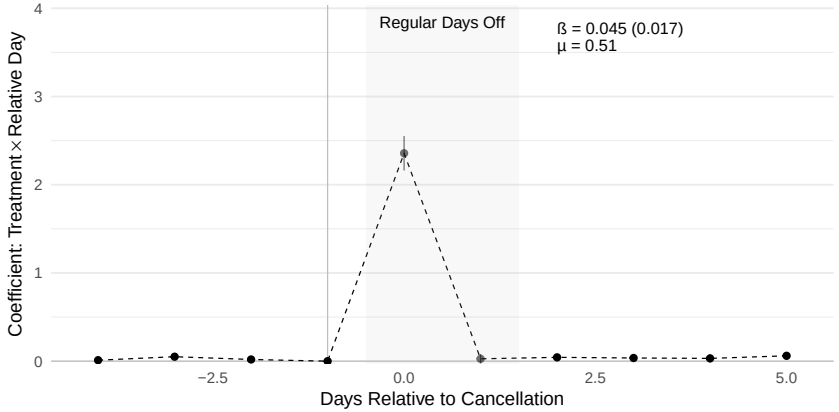
(a) Raw Data: Injury



(b) Event Study: Injury

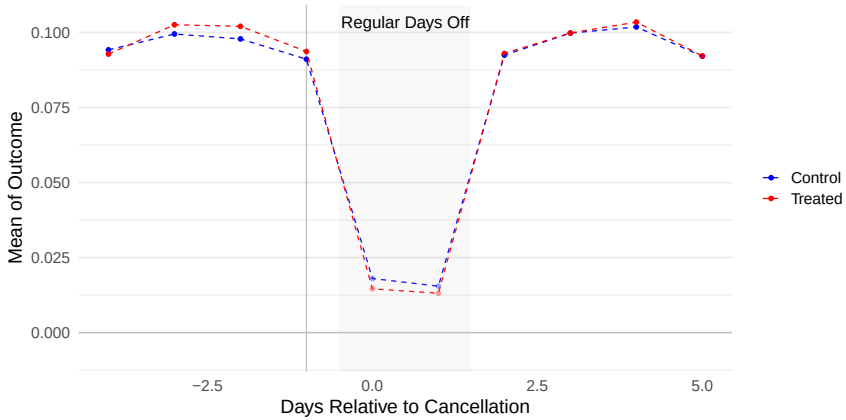


(c) Raw Data: Calls in Sick

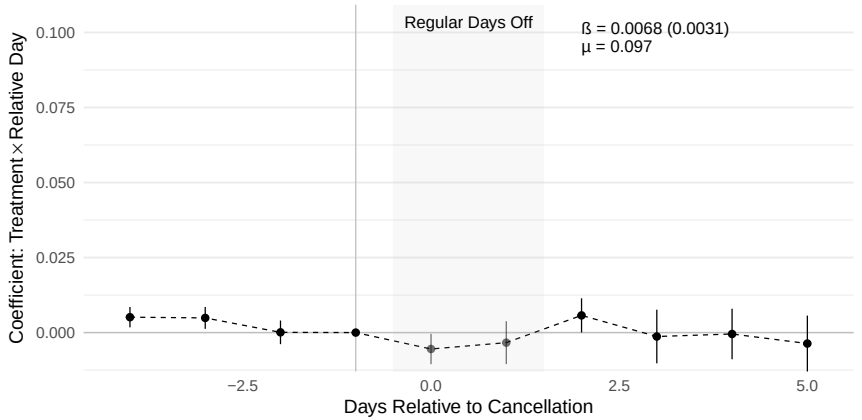


(d) Event Study: Calls in Sick

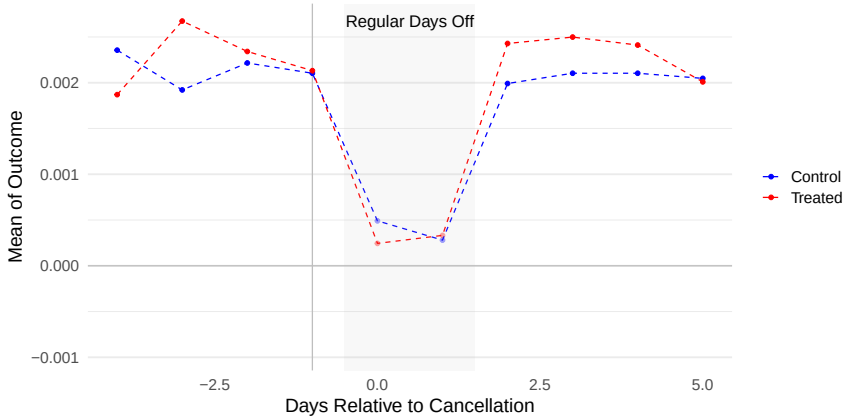
Figure C.6: *Robustness* Restricted Sample and Enforcement



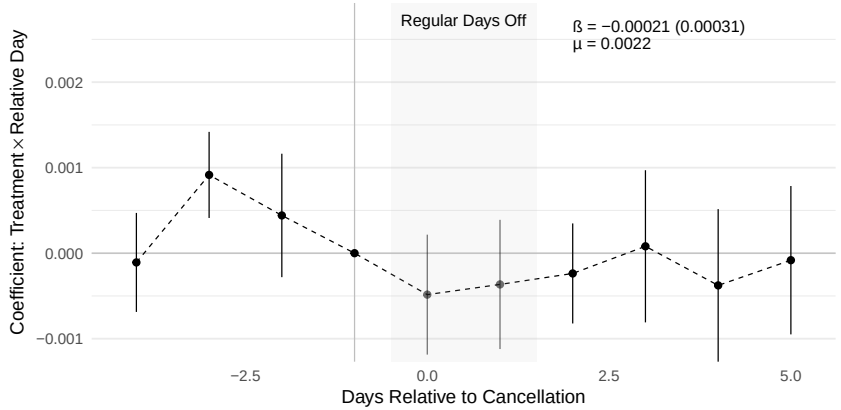
(a) Raw Data: Arrests



(b) Event Study: Arrests

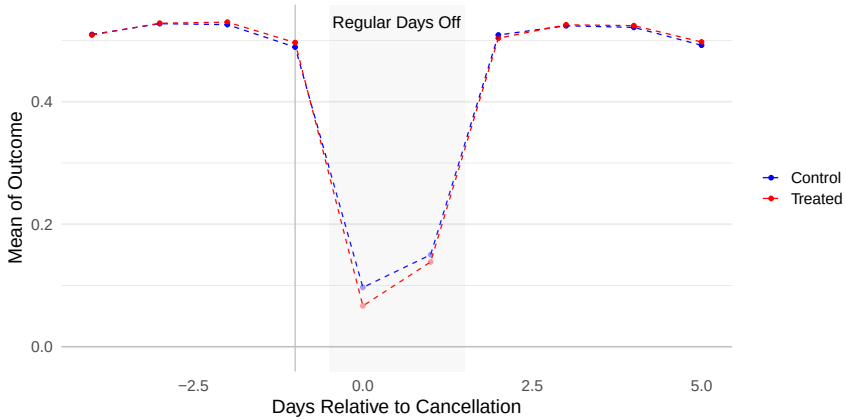


(c) Raw Data: Use of Force

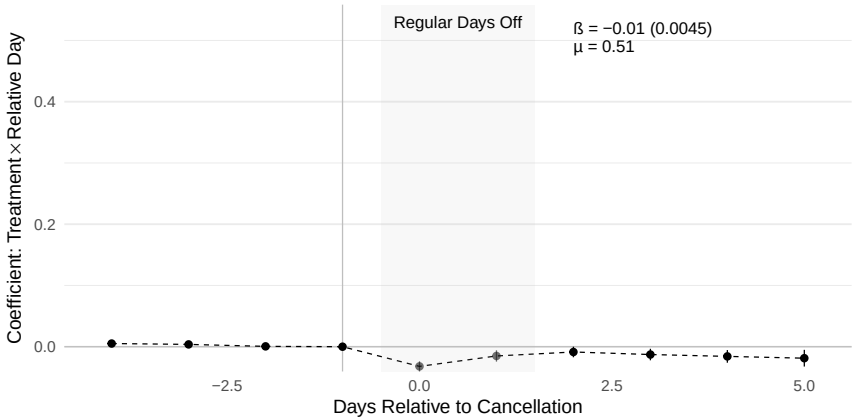


(d) Event Study: Use of Force

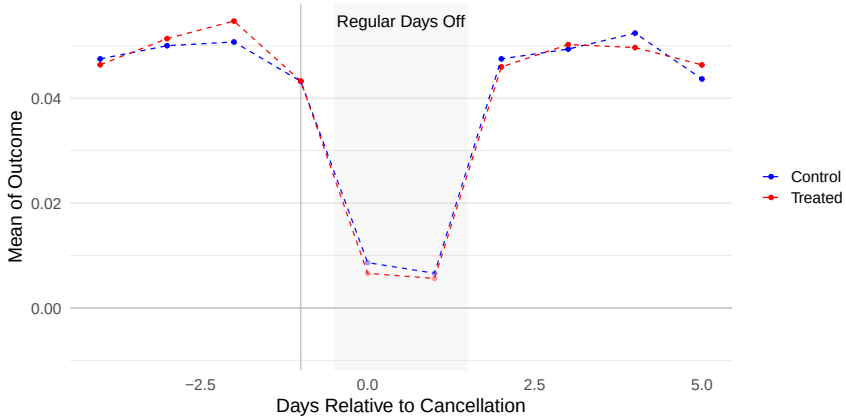
Figure C.7: *Robustness* Restricted Sample and Productivity



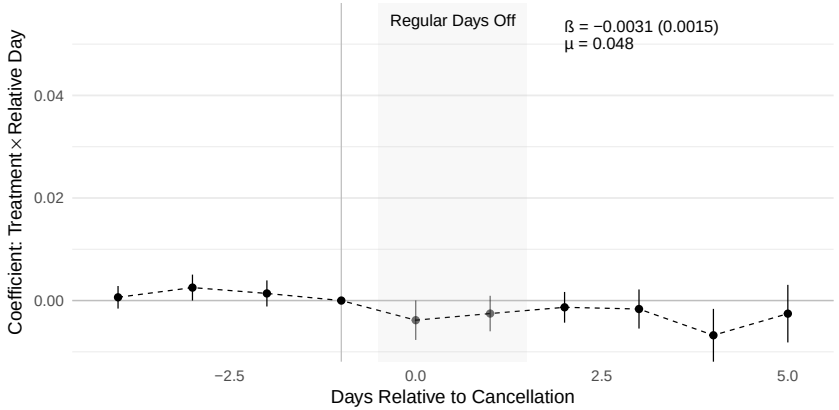
(a) Raw Data: Calls for Service



(b) Event Study: Calls for Service

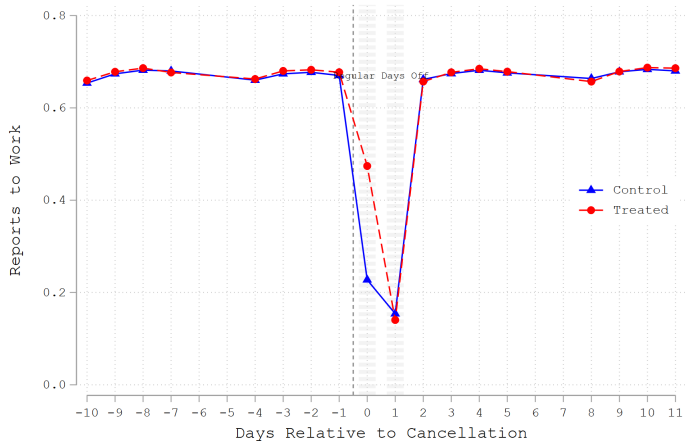


(c) Raw Data: Stops

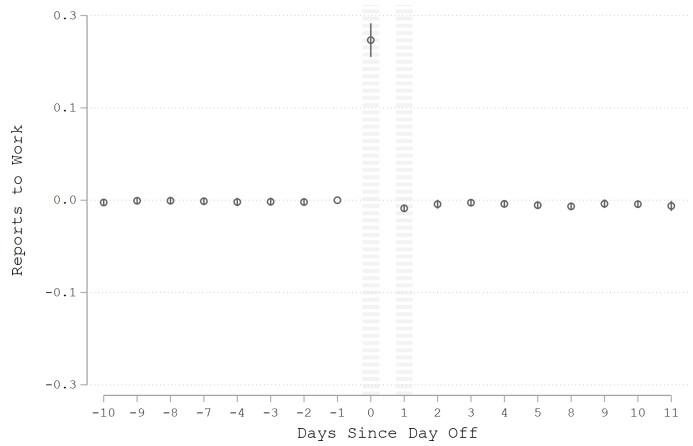


(d) Event Study: Stops

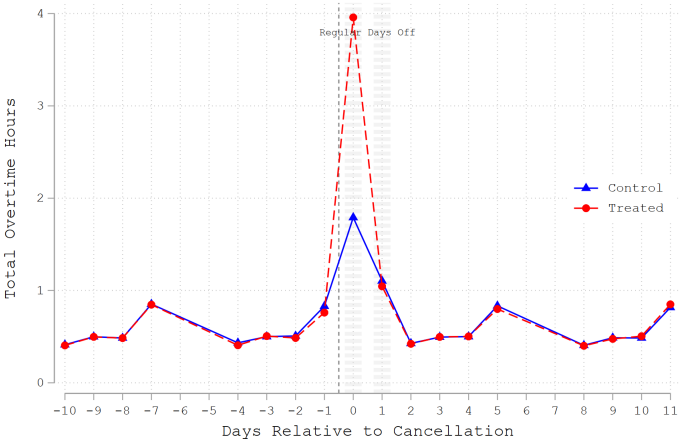
Figure C.8: *Robustness* Long Sample and Attendance



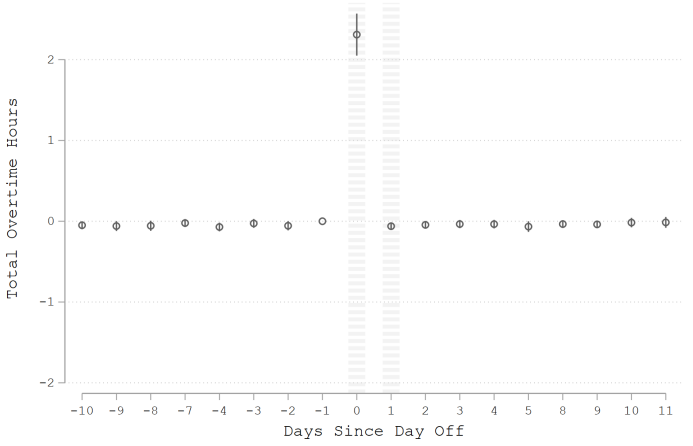
(a) Raw Data: At Work



(b) Event Study: At Work

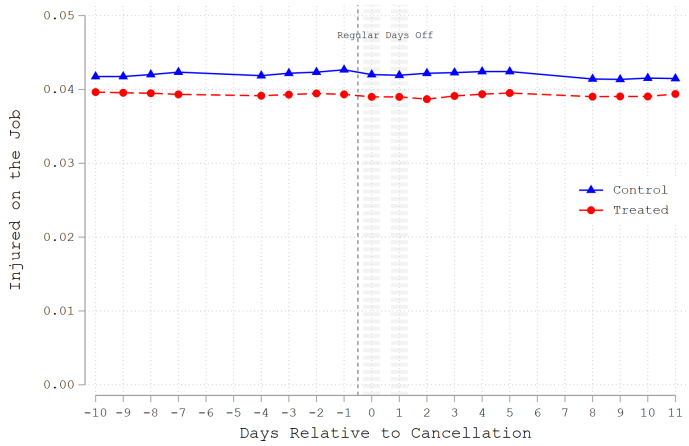


(c) Raw Data: Overtime Hours

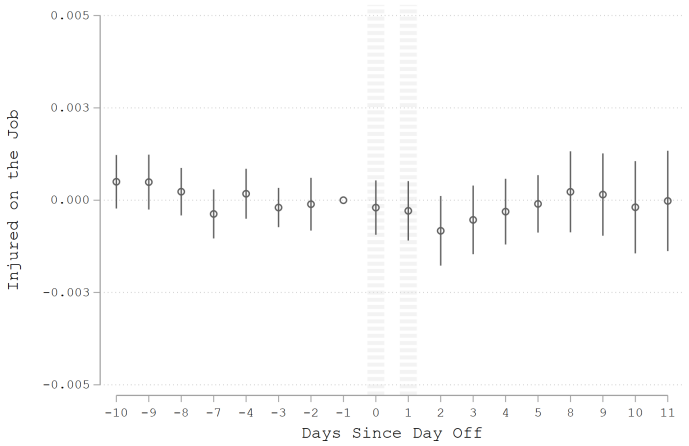


(d) Event Study: Overtime Hours

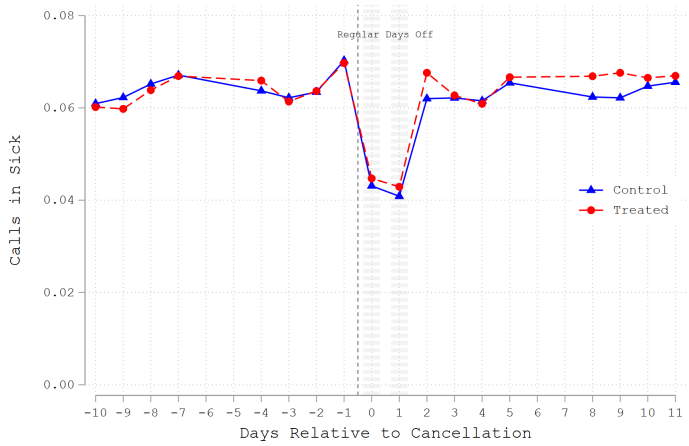
Figure C.9: *Robustness* Long Sample and Wellness



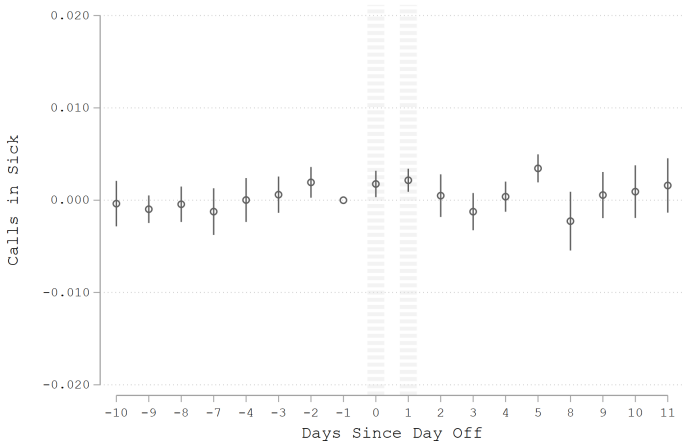
(a) Raw Data: Injury



(b) Event Study: Injury

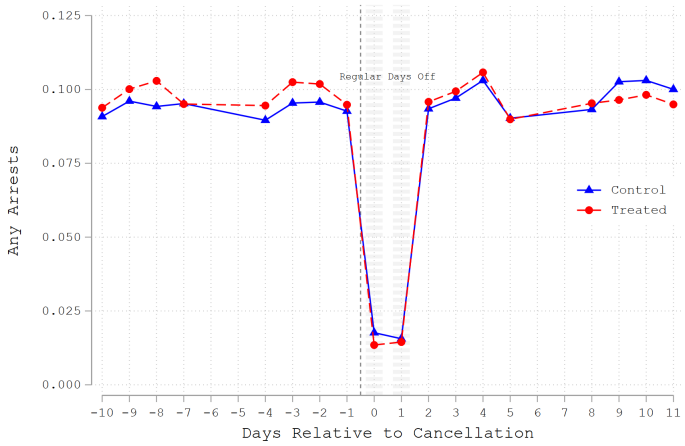


(c) Raw Data: Calls in Sick

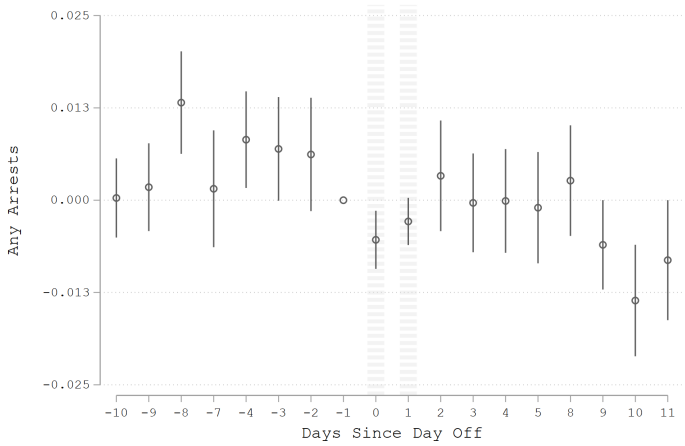


(d) Event Study: Calls in Sick

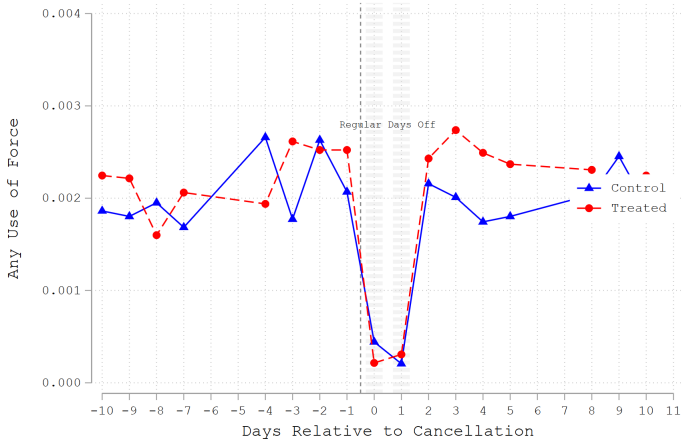
Figure C.10: *Robustness* Long Sample and Enforcement



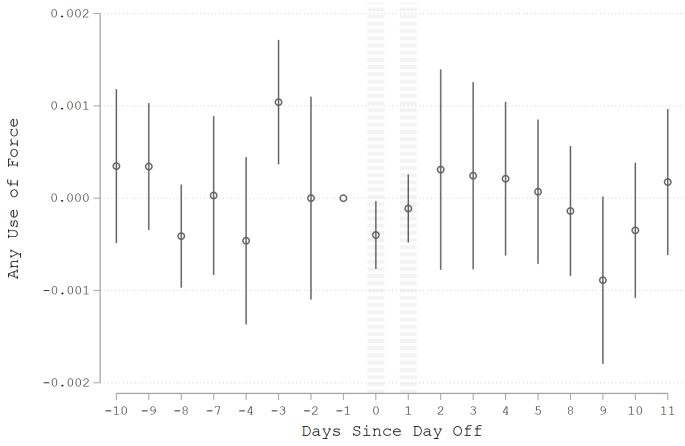
(a) Raw Data: Arrests



(b) Event Study: Arrests

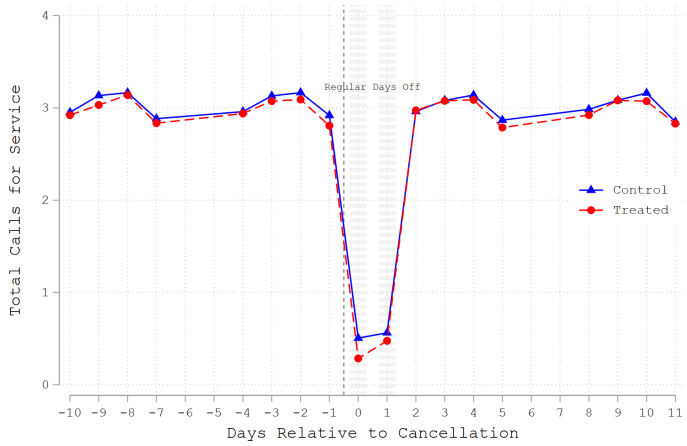


(c) Raw Data: Use of Force

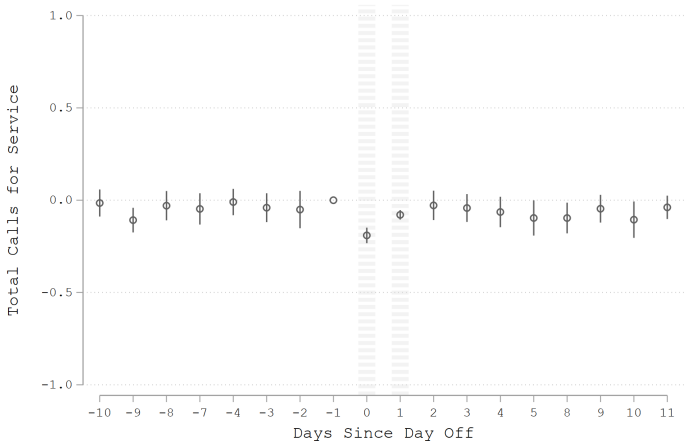


(d) Event Study: Use of Force

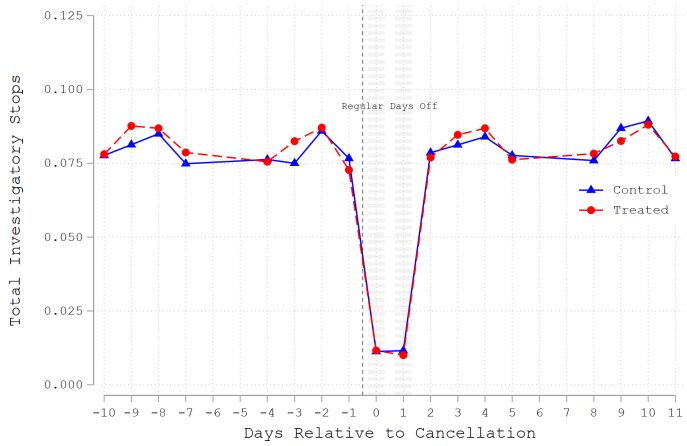
Figure C.11: *Robustness* Long Sample and Productivity



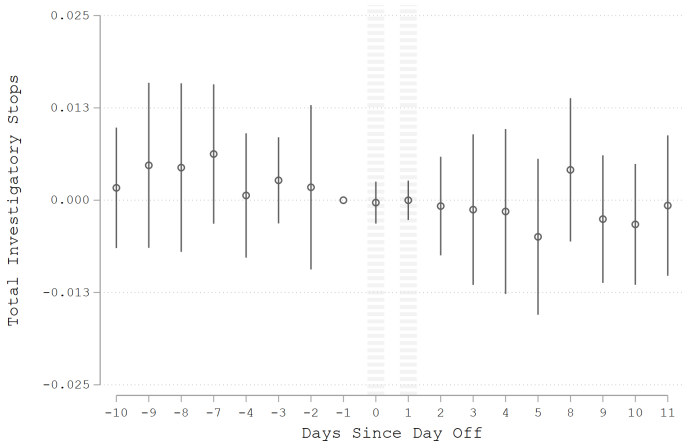
(a) Raw Data: Calls for Service



(b) Event Study: Calls for Service



(c) Raw Data: Stops



(d) Event Study: Stops