Body-worn cameras, police arrests, and racial disparities: Evidence from a large-scale staggered difference-in-differences analysis

Andrea M. Headley*
McCourt School of Public Policy, Georgetown
Ah164@georgetown.edu

Daniel B. Baker*
Department of Government and Justice Studies, Appalachian State University
bakerdb@appstate.edu

Inkyu Kang*
School of Public and International Affairs, University of Georgia
inkyu.kang@uga.edu

Keywords: Bureaucratic accountability, Body-worn cameras, Police arrests, Race, Staggered difference-in-differences

*Co-first authors with equal contribution

Abstract

Body-worn camera (BWC) programs rapidly diffused as a prominent accountability mechanism expected to improve police behavior. However, opposing theoretical predictions compete as to how BWCs will affect police actions such as arrests, and the findings from previous studies (mostly focusing on a single agency within unique contexts) remain mixed and particularly limited with regard to racial disparities. In response, this study conducts a large-scale staggered difference-in-difference analysis based on multi-year panel data on 3,360 local law enforcement agencies in the span of 10 years from 2007 to 2016. By employing Callaway and Sant'Anna's (2021) estimator, our analysis aims to provide more comprehensive and generalizable evidence of the effects of BWC adoption on police arrests in the US. Results show that BWC adoption had a negative effect on the total number of arrests, while having a null effect on the percentage of white or Black arrests out of total arrests or the ratio of white arrests to Black arrests. A major implication of our findings is that while closer oversight through BWCs led frontline officers to make arrest decisions with more caution, it alone was not enough to guard against racial disparities in arrests.

Keywords: Bureaucratic accountability, Body-worn cameras, Police arrests, Race, Staggered difference-in-differences

1 Introduction

Repeated incidents of police violence and misconduct threaten the legitimacy of police organizations. These incidents disproportionately affect residents from communities of color, leading to deficits in perceptions of police fairness and legitimacy (Edwards, Lee, & Esposito, 2019; Schwartz & Jahn, 2020) and consistent calls from Black community members for serious reform and enhanced accountability in policing (Crabtree, 2020). To counteract threats to legitimacy and faults in accountability systems, police organizations have explored a variety of strategies, including the adoption of body-worn camera (BWC) policies.

BWC policies purport to improve officer behavior, enhance the transparency of police interactions with community members, and improve accountability mechanisms, particularly those related to officer behavior and use of force (Lum, Stoltz, Koper, & Scherer, 2019; Gaub & White, 2020). Proponents of BWC policies also suggest that they may improve officer safety, lead to more accurate resolutions of complaints brought against police, and introduce training opportunities for line-level officers. These policies gained additional momentum following President Barack Obama's Task Force on 21st Century Policing, which identified promising avenues for police reform while maintaining police effectiveness (Final Report of the President's Task Force, 2015). BWCs were presented as a key policy recommendation that was supported by federal funding for agencies to adopt, purchase, and implement BWC programs.

As BWC policies proliferated, scholars examined the factors that influenced the adoption of these policies (Pyo, 2020; Nix, Todak, & Tregle, 2020), the attitudes of officers about these policies (Headley, Guerette, & Shariati, 2017; Adams & Mastracci, 2019), and the influence that these policies have on the public's perceptions of trust and legitimacy (Demir, Apel, Braga, Brunson, & Ariel, 2020; Wright & Headley, 2021). There remains mixed evidence about whether BWCs improve or alter the behavior of frontline officers. For example, some research suggests that BWCs may decrease the police use of force (Ariel, Farrar, & Sutherland, 2015; Williams, Weil, Rasich, Ludwig, Chang, & Egrari, 2021); however, there is also evidence that BWCs increase assaults against officers while failing to reduce use of force (Ariel et al., 2016). Recent reviews of BWC research from Lum et al. (2019) and Gaub and White (2020) suggest limited consensus regarding the effects of these policies due to differences in implementation and vastly different starting points in police-community relations.

This study aims to contribute to our understanding of how BWCs affect police behavior by conducting a large-scale causal study that provides relatively more comprehensive and generalizable evidence than some past research. We focus on the effect of BWC adoption on police arrests as well as racial disparities in arrests. Our study employs Callaway and Sant'Anna (2021)'s estimator for a staggered difference-in-differences analysis using multi-year panel data on 3,300 local police agencies in the US between 2007 and 2016. We find that agencies that adopted BWCs had a reduction in the total number of arrests; however, BWC adoption did not have a significant effect on racial disparities in arrests.

In the following section, we review research on factors that inform officer arrest decisions and discuss how BWC policies may be expected to influence these decisions. Within this section, we discuss monitoring and oversight mechanisms that may influence bureaucratic behavior. Next, we review the limited research on the effects of BWC policies on racial disparities in police treatment of civilians. We then discuss our data, key measures, and methods. Finally, we present results and discuss their theoretical and practical implications for considering effective accountability mechanisms in policing.

2 Review of Research

2.1 Body-Worn Camera Policies and Officer Arrest Behaviors

Police officers, in cases where arrest is not mandated by departmental policy or law, exercise immense discretion in deciding whether to arrest individuals (Huff, 2021; Walker, 1993). In many instances, officers consider legal factors when deciding to make an arrest. Legal factors include the evidence available to justify an arrest, the presence or cooperation of a victim, or the seriousness of an offense (Mastrofski, Worden, & Snipes, 1995). Research shows, for example, that offense severity, a legal factor, is a consistent predictor of arrest (Brown & Frank, 2006). In other instances, however, officers rely on extralegal factors including a person's age, sex, race/ethnicity, or the neighborhood composition (Mastrofski, Worden, & Snipes, 1995), which may undermine equity in arrest decisions. Moreover, a variety of officer-specific factors, including officer race (Brown & Frank, 2006; Mbuba, 2018), sex (Novak, Brown, & Frank, 2011), education level (Rosenfeld, Johnson, & Wright, 2020), and tenure (Mastrofski, Snipes, Parks, & Maxwell, 2000) may also be associated with arrest decisions.

Over the past decade, BWCs increased in popularity as a promising answer to calls for improved equity in outcomes of police-civilian interactions, such as arrests. BWCs are purported to be an *ex-post* accountability mechanism over bureaucratic actions, which refers to "authorities monitoring and reacting to what bureaucrats do in the performance of their jobs" (Moe, 2012, p. 2). When cameras record police-civilian encounters, these interactions are subject to potential subsequent evaluation. Research on the effects of BWCs, as well as research on accountability mechanisms in the public sector, suggest that expectations of subsequent evaluation could alter the way officers exercise discretion in police-civilian encounters. When monitoring is present, one might expect officers to be more cautious and careful when using authority and treat civilians respectfully across all interactions (Ariel et al., 2015). This aligns with public self-awareness theory, which suggests that, while under observation, individuals are more likely to comply with established rules and exhibit behaviors that align with societal norms and expectations (Duval & Wuckland, 1972; Farrar, 2013). As individual officers are more aware that their actions will – or could – come under public attention and scrutiny, officer behaviors might be more likely to align with societal expectations (Froming, Walker, & Lopyan, 1982).

Recent theoretical advancements on bureaucratic accountability in public administration also provide additional detail on the psychology of self-awareness and the behaviors of bureaucrats. The literature on felt accountability suggests that for accountability mechanisms to be effective, bureaucrats must feel the potential consequences of the accountability processes in place. In other words, bureaucrats must have a sense that they will be held accountable for their actions (Overman, Schillemans, & Grimmelikhuijsen, 2021). In implementing BWC policies, the aim of police organizations is to set the expectation that officer actions are subject to future evaluation by someone in a position of authority and that this evaluation could be tied to future sanctions or rewards in response to the officer's behavior (Hall & Ferris, 2011).

On the other hand, there are competing theoretical reasons that predict that BWCs will encourage officers to be more proactive and confident about their use of discretion and coercion (Braga, Sousa, Coldren, & Rodriguez, 2018). An initial motivation from officers for supporting the adoption of BWCs is the expectation that, in addition to protecting civilians, cameras will also protect police officers in officer-civilian interactions. This includes potential protection against physical altercations involving civilians. In addition, officers may feel more protected against misleading or spurious complaints after BWC adoption specifically because all interactions are thoroughly documented and available for review (Owens & Finn, 2018; Wallace, White, Gaub, & Todak, 2018). The video footage provides grounds that justify officer actions that would not have been available if not for BWCs. Furthermore, BWCs may lead officers to feel more obliged to take actions against people whose violation of the law is documented on camera (Ariel, 2016; Ready & Young, 2015). As street-level bureaucrats, officers often exercise discretion in ways that favor civilians based on their own moral judgments, such as by making exceptions for minor violations (Lipsky, 1980; Maynard-

Moody & Musheno, 2000). However, with BWCs in place, officers may be less comfortable routinely making exceptions or giving individuals the benefit of doubt. Instead, officers may default to a more by-the-book, legalistic approach, leading them to use a tool in their toolbelt – arrest – more than we might expect in a world prior to BWC adoption.

As these possible explanations of the effect of BWCs on officer arrest behaviors compete with one another, the body of empirical evidence also remains highly mixed and inconsistent. For instance, some studies found that BWCs led to lower levels of use of force (Ariel et al., 2015; Braga, Coldren, Sousa, Rodriguez, & Alper, 2017) and fewer arrests (MacDonald, Fagan, & Geller, 2016; Braga, Barao, Zimmerman, Douglas, & Sheppard, 2020). By contrast, other studies found that officers equipped with BWCs made more arrests and issued more citations compared to a non-BWC control group (Braga et al., 2017). Meanwhile, a fair number of studies found null effects of BWCs on police behavior, supporting neither of the anticipated effects on officer behaviors. For instance, Hedberg, Katz, and Choate (2017) and Ready and Young (2015) found that BWCs did not have a significant influence on officer arrest behaviors. In a similar vein, Grossmith, Owens, Finn, Mann, Davies, and Baika (2015) and Headley et al. (2017) found no effects on the routine, discretionary officer activities of stop-and-frisk and traffic stop decisions of officers, respectively. Yokum, Ravishankar, and Coppock (2019) also found no significant effects of BWCs on officer use of force.

The motivation driving agencies to adopt BWCs hinges on a number of political as well as bureaucratic factors such as organizational leadership, culture, or even local political atmosphere. For example, some police agencies might have a leader who pushes for officer performativity and increasing public safety outcomes, whereas other agencies might be facing stronger external demands from politicians and stakeholders to prioritize procedural values in policing. The way agencies write details of the BWC policy, such as how the device should be used or how recorded video footage should be disclosed (Lum et al., 2017), will vary a great deal depending on how they intend to utilize

BWCs, which will cause variations in the way officers change behaviors once the cameras are assigned. Relatedly, it is worth noting that a majority of existing studies on the relationship between BWCs and police behavior have been situated in a single police agency – while such evidence may be meaningful within the situated context it may not necessarily be generalizable to many other agencies in the US.

2.2 Body-Worn Camera Policies and Racial Disparities in Arrests

Racial disparities abound in policing. Black individuals are more likely to be stopped, searched, frisked, and arrested compared to their white counterparts (e.g., Headley, 2020, Headley & Wright, 2020). With regard to arrests specifically, early research demonstrated that Black people were arrested at higher rates; however, the reasons attributed for these disparities varied (Black, 1971; Smith & Visher, 1981). For instance, using systematic social observations, Black (1971) argued it was related to the demeanor of Black individuals when interacting with the police, whereas Lundman and colleagues (1978) noted that Black people were more likely to request formal action for Black youth. More recent research has confirmed earlier work on Black suspects being more likely to be arrested than white suspects (Brown & Frank, 2007; Kochel, Wilson & Mastrofski, 2011), whereas Smith, Visher and Davidson (1984) found that the race of the victim or complainant matters for arrest decisions, so police may be more likely to make an arrest when there is a white complainant.

Despite racial justice being central to the diffusion of BWC programs and the calls for police accountability, studies have cast doubt that even though BWCs enable closer monitoring and oversight over police actions they are unlikely to provide effective solutions for racial equity in policing. The doubts are grounded on the idea that racial bias is largely a matter of human intent that is difficult to prove even with video footage (Birzer & Birzer, 2006; Coleman & Kocher, 2019; Garrett, 2000). BWCs exist as a monitoring mechanism of observable police behavior when in the field; however, even with the video footage produced by BWCs, whether a police behavior is driven by implicit biases and prejudice or commitment to legitimate law enforcement is difficult to tease out and evaluate objectively. Based on a comprehensive review of the current practice of BWC usage in large metropolitan cities, Murphy (2019) stated: "Though body-worn cameras may go some way to curbing racial bias in policing, they are almost certainly not the panacea that some people initially hoped they would be. In order to best reduce racial bias in policing, what is needed is change in recruiting, organizational culture, training, and accountability mechanisms." (p. 182).

There is consistent discussion of BWC policies as an essential path to enhancing officer accountability. The expectation is that this enhanced accountability through officer oversight will help confront racial injustice at the frontline. However, the empirical evidence on the link between BWC adoption and racial disparities in police actions is scarce. Even the few studies that looked at the issue offered conclusions that contradict the effectiveness of BWCs for improving racial justice in frontline policing. For example, Huff (2021) suggested that higher percentages of Black individuals in a neighborhood is significantly associated with an increase in arrests, and the arrests were more likely to occur in both Hispanic and Black neighborhoods before *and* after BWC deployment (Huff, 2021), indicating that BWC programs might not be effective in reducing racial disparities in police actions as policy expects had hoped. This study seeks to examine these expectations across a variety of police departments in the U.S. in the wake of large-scale adoption of BWC policies.

3 Methods

3.1 Data

We combined two sources of data to create an agency-level multi-year panel dataset: 1) Uniform Crime Reporting (UCR) program data, and 2) Law Enforcement Management and Administrative Statistics (LEMAS) survey 2016 data. UCR program is the most comprehensive, publicly available nationwide data on police arrest and other basic agency-level characteristics in the US. LEMAS surveys have been conducted periodically by the Bureau of Justice Statistics since 1987. The survey involves a nationally representative sample of thousands of general-purpose law enforcement agencies each year, including local and county police, sheriffs' offices, and primary state police and highway patrol units[1]. Agencies with more than a hundred full-time sworn personnel are reached with certainty, while smaller agencies with less than a hundred full-time sworn personnel are sampled through strata sampling based on the number of full-time sworn personnel and type of agency.

3.2 Treatment and Pre-Treatment Covariates

In 2016, the LEMAS BWC supplementary project contained data on whether and when agencies adopted a BWC program. The survey first asked each agency "*When did your agency first get body-worn cameras?"* and then asked a follow-up question for the agencies that did acquire BWCs, "*How would you describe the current state of body-worn camera deployment in your agency?*". Answers to the second question fell under one of four categories: (1) exploratory/pilot deployment, (2) partial deployment, (3) complete deployment for some assignments/partial deployment in others, or (4) full deployment to all intended personnel. Among 3,899 agencies that responded to the first question, 1,900 agencies said they have acquired BWCs while 1,999 agencies said they have not. Among 1,900 agencies that said they have acquired BWCs, 1,839 agencies responded to the follow-up question about the current state of BWC deployment. First, 443 agencies said their program was an exploratory/pilot deployment in the 2016 LEMAS survey. These agencies were excluded from our analysis since they never moved on to a formal program adoption as of 2016 and count as neither treated units nor non-treated units. Among the rest of 1,396 agencies that reached a partial or a full deployment of BWCs, 22 agencies that acquired cameras in 2007 or before (as early as 1994) were dropped as we confined our time frame to

[1] This excludes Hawaii, which does not have a primary state police force.

between 2007 and 2016 (2007 was used as pre-treatment period for units treated in 2008). The purpose was to keep a reasonable number of time points while limiting major extraneous events that may cause history bias, which in this case were two events that took place in 2008: The Great Recession and the Presidential election that put the Obama Administration in power.

As a result, we had 1,999 agencies in the comparison group (i.e., "never-treated" group) and 1,396 agencies in the treatment group (i.e., agencies that adopted a formal BWC program at some point between 2008 and 2016), which added up to 3,395 agencies in total. Given the focus of our analysis at local police authorities, 35 state highway patrol agencies were excluded (4 from the treatment group and 31 from the comparison group). Therefore, our final sample size was 3,360 observations per year, adding up to 33,600 from 2007 to 2016 (N=33,600).

Although it is difficult to empirically verify parallel trends in a difference-in-differences approach, incorporating pre-treatment covariates helps mitigate the concerns for dynamic selection bias; a situation where differences in pre-treatment unit characteristics are correlated with treatment selection and further affect the evolution of outcomes (Houngbedji, 2016; Abadie, 2005), as in the famous case of "Ashenfelter's dip" (Ashenfelter, 1978). For example, there could have been some state-level factors, such as a unique political atmosphere, that affected both the adoption of the BWC program among local police agencies and the arrests made by the agencies. Furthermore, agency type and size, which are the basic covariates in organizational-level management studies, could also have affected both the BWC adoption and the arrests made by the agencies. To mitigate these concerns, we utilized state (excluding Hawaii; see footnote 2 for reference) and agency size/type as two pre-treatment covariates to estimate the propensity score and the following outcome regressions. Table 1 summarizes the basic pre-treatment covariates and the treatment onset for the treatment group and the comparison group.

Table 1 Descriptive Statistics

	Treatment group		Comparison group (never-treated group)	
Pre-Treatment Covariates				
Agency Type and Size				
Sheriff's offices with 100 or more full-time sworn officers	207	14.87%	234	11.89%
Sheriff's offices with 10-99 full-time sworn officers	415	29.81%	680	34.55%
Sheriff's offices with less than 10 full-time sworn officers	453	32.54%	561	28.51%
Local and county police with 100 or more full-time sworn officers	87	6.25%	150	7.62%
Local and county police with 10-99 full-time sworn officers	164	11.78%	244	12.40%
Local and county police with less than 10 full-time sworn officers	66	4.74%	99	5.03%
*Region (broken down into states in the analysis)**				
Northeast	109	8.14%	528	27.95%
West	254	18.97%	196	10.38%
Midwest	354	26.44%	636	33.67%
South	622	46.45%	529	28.00%
BWC Adoption Year				
2008	20	1.40%	N/A	
2009	37	2.60%		
2010	71	4.98%		
2011	60	4.21%		
2012	88	6.18%		
2013	113	7.93%		
2014	301	21.12%		
2015	491	34.46%		
2016	244	17.12%		
Observations Per Year	1,392		1,968	

*States were not consolidated into census regions in the actual analyses.

Our dependent variables fell under one of two categories: (1) total arrest and (2) percentage of arrest by race. First, total arrest was measured by the number of arrests an agency made in a given year. Second, the percentage of arrests by race was measured by the number of white arrests out of total arrests, the number of Black arrests out of total arrests, and the ratio of white arrests to Black arrests. In sum, our analysis had four dependent variables.

3.3 Estimation Strategy

We employed the Callaway and Sant'Anna (2021) estimator to estimate the effect of BWC adoption on the outcome variables, using *csdid* command in Stata 17. Arguably, one of the most common approaches to identifying treatment effects using panel data is difference-in-differences (DiD). Under a parallel-trend assumption, DiD produces an unbiased estimate of the Average Treatment Effect on the Treated (ATT). When the intervention occurred in multiple time periods, the model diverges from the canonical two-group/two-period DiD (i.e., treatment group and comparison group/pre- and post-intervention period) where Two-Way Fixed Effects (TWFE) regression is typically estimated. According to De Chaisemartin and d'Haultfoeuille (2020), roughly 20% of all empirical studies that were published in the *American Economic Review* between 2010 and 2012 administered a TWFE regression to estimate the effect of an intervention on an outcome. The idea of TWFE regression is to adjust for time-invariant heterogeneities across different units as well as unit-invariant heterogeneities across different time points. Consider a case where time points are denoted by t where $t = 1, \ldots, T$. To estimate the effect of a treatment that occurred in points in time, TWFE linear regression is typically estimated which can be written as follows:

$$Y_{it} = \theta_t + \eta_i + \beta D_{it} + v_{it}$$

where θ_t is a time fixed effect, η_i is a unit fixed effect, D_{it} is a dummy variable for treatment, v_{it} contains time-varying unobserved heterogeneities, and β is the parameter of interest. Although β is often interpreted as the direct representation of the ATT, it is in fact a "weighted average of all possible two-group/two-period DiD estimators in the data" (Goodman-Bacon, 2021, p. 254).

For TWFE estimators to be unbiased – namely, for all possible two-group/two period DID designs to be "good" designs – the treatment effect must have been constant across groups as well as over time. That is, there must have been no heterogeneous treatment effects and also no dynamic effects. This assumption is bold and is unlikely to hold in a majority of applied settings including the present study: "When the treatment effect is constant across groups and over time, [TWFE] regressions identify that effect under the standard "common trends" assumption. However, it is often implausible that the treatment effect is constant" (De Chaisemartin & d'Haultfoeuille, 2020, p. 2964). This potential pitfall of TWFE has attracted attention in the recent literature on econometrics and program evaluation (e.g., Baker, Larcker, & Wang, 2022; Sun & Abraham, 2021). In the present study, if we were to apply traditional TWFE regression, we must assume that the effects of BWC program on arrest variables were identical across agencies of different types (e.g., local police versus sheriffs' offices), different size, different location, and so on. We must also assume that once the program was implemented, its effects on arrest variables remained the same over time. These assumptions are unlikely to hold from a logical standpoint.

In an attempt to overcome these limitations of TWFE regression, Callaway and Sant'Anna (2021) recently proposed a difference-in-difference estimator for situations where treatment occurred in multiple time periods but effect heterogeneity and/or dynamic effect are anticipated. Their strategy is essentially to break down the single DiD into many different pieces of two-by-two DiDs and discard the bad two-by-two DiDs where treated units are included in the comparison group. They propose a nonparametric point-identification of what they refer to as the "group-time average treatment effect"

– namely, the average effect for group *g* at time *t*, where a "group" refers to a cohort that first received the treatment at a given time point. Assuming a "staggered" treatment adoption – i.e., once a unit participates in the treatment, it is irreversible and cannot be turned off – which is the case of the present study where agencies never switch from a treated unit to a non-treated unit, Callaway and Sant'Anna (2021) set up a potential outcomes framework as follows:

$$Y_{i,t} = Y_{i,t}\ (0) + \sum_{g=2}^{T} \{Y_{i,t}\ (g) - Y_{i,t}\ (0)\} \cdot G_{i,g}$$

Where $Y_{i,t}\ (0)$ denotes unit *i*'s untreated potential outcome at time *t* if the unit never receives the treatment throughout the available time periods, $Y_{i,t}\ (g)$ denotes the potential outcome of unit *i* at time *t* if they were to first receive the treatment in period *g* for $g = 2, \ldots, T$, G_g denotes a binary variable that equals one if a unit first receives the treatment in period *g*. Since different potential outcomes for the same unit at the same period can never be observed simultaneously, group-time average treatment effect can be expressed as follows:

$$\textit{Group-Time Average Treatment Effect} = ATT\ (g, t) = E[Y_t\ (g) - Y_t\ (0) \mid G_g = 1], \text{ for } t \geq g$$

Potentially a large number of group-time average treatment effects are aggregated to provide a general-purpose summary parameter that is analogous to ATT in two-group/two-period DiD. The default choice of estimand for parametric estimation of the group-time average treatment effect in the *csdid* package is doubly robust estimand based upon inverse probability of tilting and weighted least squares proposed by Sant'Anna and Zhao (2020), which we implemented in our analysis.

An additional advantage with Callaway and Sant'Anna (2021) estimator is that it allows researchers to easily incorporate pre-treatment characteristics and relax parallel trend assumption to "conditional" parallel trend assumption. Any pre-treatment characteristics can be adjusted for so long as they are not unique to either treatment group or comparison group at any given time period. As previously discussed (see Table 1 for reference), we used two pre-treatment covariates to relax the parallel trend assumption: 1) state in which each agency was located and 2) agency type/size (i.e., local or county police departments versus sheriffs' offices with more or less than 100 full-time sworn officers). As such, we can rule out the dynamic selection bias in scenarios where the state and/or the size and type of an agency affects both the adoption of BWCs and the arrests made by the agency.

4 Results

4.1 Estimated Effects of Body-worn Camera Adoption on Police Arrest

We present the aggregate group-time average treatment effect of the BWC adoption on the specified dependent variables. First, the adoption of BWCs had a significant negative effect on the total number of arrests made by agencies (Coef.=-193.35, *p*-value=0.046). To get a better sense of the variations in the effect of BWC adoption on total arrests across different treatment onsets, we also estimated the effects by calendar year. As it turns out, the negative effect of BWC adoption on total arrests at a given year was statistically significant in 2016 at the significance level of 0.05 (Coef. -305.20, *p*-value=0.002). Next, we find little evidence that the BWC adoption made any difference in racial disparities in arrest. It turns out that the program adoption did not have any significant effect on either the percentage of white arrests (Coef.=0.001, *p*-value=0.549), the percentage of Black arrests (Coef.=-0.001, *p*-value=0.548), or the ratio of white arrests to Black arrests.

Overall, our results were not sensitive to the use of pre-treatment covariates (i.e., state and agency type/size). We ran the identical analysis without the covariates and found that the statistical significance of the effects of BWC adoption on the dependent variables remained the same, except for the number of Black arrests. For this outcome measure, BWC adoption turned out to be significant at the significance level of 0.05 (Coef. -105.78, *p*-value=0.048). We summarize these sensitivity analysis results in Table A1 in the Appendix.

Table 2 Effects of BWC Adoption on Police Arrest (Aggregated ATT)

Dependent Variables	Coef.	S.E.	z	P>|z|	n†
Total Arrest					
Number of arrests	-193.35	96.90	-2.00	0.046**	24,400
Number of white arrests	-108.69	56.68	-1.92	0.055	24,357
Number of Black arrests	-103.63	53.40	-1.94	0.052	21,145
Racial Disparities in Arrest					
Percentage of white arrests	0.001	0.002	0.64	0.525	24,357
Percentage of Black arrests	-0.001	0.002	-0.62	0.536	21,145
White arrests to Black arrests ratio	-0.548	1.061	-0.52	0.606	21,105
Year = 2007 to 2016, N = 33,600 (3,360 observations × 10 years).					

† Estimations were performed using observations with pairs balanced at pre-treatment and post-treatment periods.
$p < 0.05$: **, $p < 0.01$: ***, two-tailed test.

Table 3 Effects of BWC Adoption on the Number of Arrests by Treated Year

Effects on Total Arrest	Coef.	S.E.	z	P>|z|
Effects by Year				
Effect in 2008	-167.16	135.33	-1.24	0.217
Effect in 2009	-108.65	87.71	-1.24	0.215
Effect in 2010	-103.43	76.22	-1.36	0.175
Effect in 2011	-201.88	149.69	-1.35	0.177
Effect in 2012	-208.89	138.71	-1.51	0.132
Effect in 2013	-210.85	160.82	-1.31	0.19

Effect in 2014	-253.71	137.00	-1.85	0.064
Effect in 2015	-30.23	89.45	-0.34	0.735
Effect in 2016	-304.04	97.18	-3.13	0.002**
Year = 2007 to 2016.				

p < 0.05: **, *p* < 0.01: ***, two-tailed test.

4.2 Event Study Effects and Parallel Trend Assumption

To check on whether the parallel trend assumption was met, we also estimated the dynamic event study effects based on Callaway and Sant'Anna (2021). There is no guarantee that the estimated pre-trend event study effects would have remained the same in the post-intervention period. Nevertheless, they still provide useful information as to whether the treatment group and the control group were comparable in the pre-treatment period which lends support to the parallel trend assumption (Li, 2022).

It turns out that the pre-trend was negative and statistically significant for eight years before the BWC adoption (Coef.=-58.50, *p*-value=0.007). But the pre-trend was statistically indistinguishable from zero for three years prior to the BWC adoption (Coef.=-38.17, *p*-value=0.087). The fact that our conditional parallel trend assumption held for three years before treatment, which we believe is a reasonable length of time, lends support for us to infer causality of sorts between BWC adoption and the specified dependent variables from our results. The event study effects for the time frame between three years before treatment and eight years after treatment are summarized in Table 4 and visualized in Figure 1.

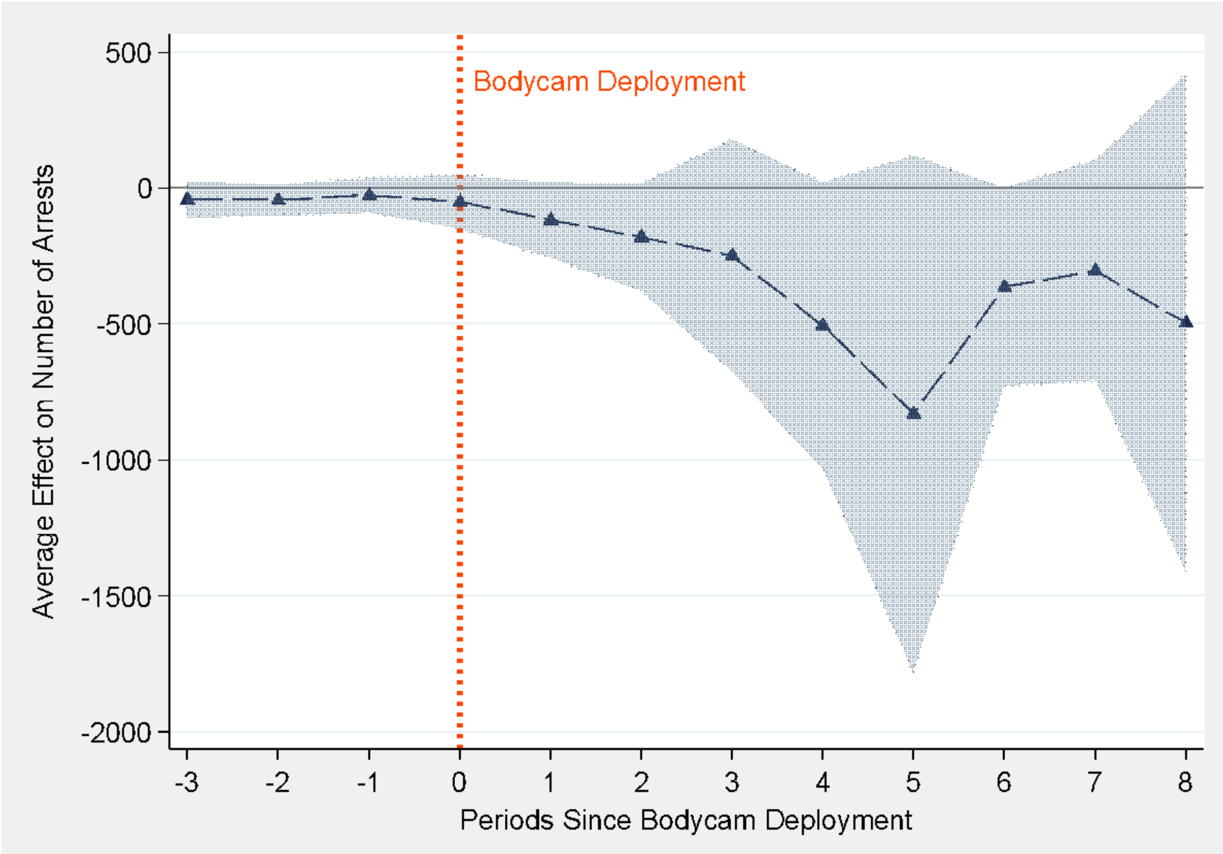
Table 4 Event Study Effects

	Coef.	**S.E.**	**z**	**P>|z|**
Pre-Trend Average	-38.17	22.32	-1.71	0.087
Post-Trend Average	-344.58	152.86	-2.25	0.024**

3 Years Before Treatment	-42.53	34.11	-1.25	0.212
2 Years Before Treatment	-45.10	28.87	-1.56	0.118
1 Years Before Treatment	-26.88	32.46	-0.83	0.408
Treatment Onset	-50.71	50.93	-1	0.319
1 Years After Treatment	-118.14	71.37	-1.66	0.098
2 Years After Treatment	-180.43	101.12	-1.78	0.074
3 Years After Treatment	-249.21	217.96	-1.14	0.253
4 Years After Treatment	-506.51	269.44	-1.88	0.06
5 Years After Treatment	-831.17	485.00	-1.71	0.087
6 Years After Treatment	-365.04	185.88	-1.96	0.05
7 Years After Treatment	-303.86	207.27	-1.47	0.143
8 Years After Treatment	-496.19	467.11	-1.06	0.288
Year = 2007 to 2016.				

$p < 0.05$: **, $p < 0.01$: ***, two-tailed test.

Figure 1 Dynamic Event Study Effects



5 Discussion and Conclusion

BWC policies have been proposed by communities calling for police reform, and their adoption has rapidly expanded across police forces in the US in recent years (Huff, 2021). BWCs provide a better way to monitor and review police-civilian interactions (Lum et al., 2019), which should ultimately improve police behavior and enhance accountability. However, competing theoretical predictions on how BWCs will impact police behavior coexist, and the mixed and inconclusive evidence from previous empirical studies have focused on a single agency within a unique context. In response, the purpose of this study was to acquire more generalizable knowledge as to how BWCs affected police behavior in the US by conducting a large-scale staggered difference-in-difference analysis using multi-year panel data on thousands of local-level police authorities. With a focus on arrest decisions, our analysis targeted two main points. First, we examined the effects of BWC adoption on the total number of arrests in a given year. In addition, we analyzed BWC adoption on changes in racial disparities in arrest (number, percent and ratio of Black and White arrests).

Our results suggest that BWC adoption led to a reduction in the total number of arrests. It shows that overall, video-recording police-civilian encounters escalated pressures for officers to be more careful and cautious when making arrest decisions. That is, BWCs generally served as a control mechanism rather than a protection for police officers. This should not be surprising given that the diffusion of BWC programs in the US was embedded in the societal context where high-profile police killings and brutality, especially against minorities, fueled movements for procedural accountability in policing. Even though a few existing studies focusing on a single agency showed that BWCs may increase police arrest (e.g., Braga et al., 2017), such as by providing protection for officers from off-based complaints, our results suggest that those studies may not be generalizable to other parts of the US. It is not uncommon that researchers encounter mixed results from multiple impact analyses on the same interventions and outcomes, not just on BWCs and police behavior but also on other policy issues. One major reason for this is that the impact of public policies hinges on what bureaucratic agencies do and why they do it while implementing the policies. Yet, impact analyses often neglect bureaucratic factors that play a critical role during the implementation phase, despite being methodologically rigorous. As Heineman, Bluhm, Peterson, and Dearny (1990) argued, "The economics background common to many analysts...can facilitate narrow, technically proficient analysis of a problem without adequate regard for the administrative or political process" (p. 63). We believe that there are values to combining lessons from the literature on bureaucracy and implementation to guide and contextualize policy evaluation studies.

Our results also found very little evidence that BWC adoption impacted racial disparities in police arrests. Specifically, BWC adoption did not have a significant effect on the percentage of white arrests, percentage of black arrests, or ratio of white arrests to black arrests. There are numerous reasons that could potentially explain the null results with regard to racial disparities. Racial disparities in police behavior are well documented in the literature (e.g., Headley, 2020). There are many identified potential sources of disparities, including interpersonal biases (Glaser, 2015) as well as institutional policies and practices (Epp et al., 2014). If disparities in arrest come from subjective bias by police officers, BWCs will not be entirely useful since subjective internal decision-making processes will not be caught on camera, as previously discussed. Only the behavior of the officer and the subsequent outcome of the encounter will be objectively recorded and made available for review. Thus, unless racism is blatantly obvious, as is the case sometimes, then even video recordings cannot fully eliminate information asymmetry when it comes to bureaucratic biases and prejudice that are subjective by nature. Even with BWCs, it would still be difficult to tease out and judge whether a police measure was driven by racial bias or commitment to legitimate law enforcement. By contrast, if the racial disparities come from institutional policies or practices that encourage inequitable use of force, law enforcement, or distribution of resources that marginalize minority communities, BWCs would still not be able to correct for these issues either, as officers would still be subject to the same policy or performance incentives. In other words, whether the racial injustice in police actions is caused by subjective biases of individual officers or by flawed organizational-level policies and practices, enhanced oversight through BWCs alone will unlikely address the full extent of the problem.

Our study is not without limitations. With regard to data and measurement, there are three primary limitations that could impact the interpretation and confidence in the findings. First, with regard to measuring racial disparities in arrests, there is a larger literature critiquing alternative approaches to measurement and appropriate benchmarks (Knox & Mummolo, 2020). This is particularly important when drawing conclusions about biased or discriminatory police behavior (Gaebler et al. 2020; Neil & Winship, 2019; Tregle, Nix & Alpert, 2019). Thus, any conclusions about the link between effects of BWCs on racial bias must be drawn with caution from our findings, as we are neither attributing motive nor causality herein. Second, our organizational-level quasi-experimental approach does not allow us to unpack specific mechanisms between BWC adoption and police arrest decisions. We believe that the literature will immensely benefit from more qualitative or mixed-methods studies that explore nuanced perspectives of individual officers about the subject matter. Finally, we are limited in the conclusions we can draw about the entire population of municipal agencies in the US due to issues regarding missing data and sample size considerations. Apart from representativeness of the sample, data is rarely ever missing at random and, as such, poses challenges for statistical power and unbiased parameter estimates.

Limitations notwithstanding, this study provides more generalizable evidence on the effects of BWCs on police arrest by analyzing thousands of agencies over the span of 10 years, which complements previous studies that were predominantly focused on a single agency within a unique context. It also suggests theoretical implications for how accountability mechanisms that apply cutting-edge technology can impact the output of street-level bureaucrats. Our findings echo other examples of the limited equity effects of monitoring and oversight on improving bureaucratic behavior at the frontline, identifying that BWCs alone are not enough to correct racial disparities in police arrests. Instead, there must be a concerted managerial effort beyond accountability mechanisms, such as cultural reform, leadership changes, or training and education efforts, to address the fundamental problems in policing and public service at the frontline.

6 References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The review of economic studies*, *72*(1), 1-19.

Adams, I., & Mastracci, S. (2019). Police body-worn cameras: Effects on officers' burnout and perceived organizational support. *Police quarterly*, *22*(1), 5-30.

Ariel, B., Farrar, W. A., & Sutherland, A. (2015). The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial. *Journal of quantitative criminology*, *31*(3), 509-535.

Ariel, B., Sutherland, A., Henstock, D., Young, J., Drover, P., Sykes, J., ... & Henderson, R. (2016). Wearing body cameras increases assaults against officers and does not reduce police use of force: Results from a global multi-site experiment. *European journal of criminology*, *13*(6), 744-755.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 47-57.

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates?. *Journal of Financial Economics*, *144*(2), 370-395.

Birzer, M. L., & Birzer, G. H. (2006). Race matters: A critical look at racial profiling, it's a matter for the courts. Journal of Criminal Justice, 34(6), 643-651.

Black, D. (1971). "The social organization of arrest", Stanford Law Review, Vol. 23, pp. 1087-1111.

Braga, A. A., Barao, L. M., Zimmerman, G. M., Douglas, S., & Sheppard, K. (2020). Measuring the direct and spillover effects of body worn cameras on the civility of police–citizen encounters and police work activities. *Journal of Quantitative Criminology*, *36*(4), 851-876.

Braga, A. A., Coldren, J., Sousa, W., Rodriguez, D., & Alper, O. (2017). The Las Vegas body-worn camera experiment: research summary. *Annotation*.

Brown, R. A., & Frank, J. (2006). Race and officer decision making: Examining differences in arrest outcomes between black and white officers. *Justice quarterly*, *23*(1), 96-126.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200-230.

Coleman, M., & Kocher, A. (2019). Rethinking the "gold standard" of racial profiling: § 287 (g), secure communities and racially discrepant police power. American Behavioral Scientist, 63(9), 1185-1220.

Crabtree, S. (2020). Most Americans say policing needs 'major changes,'. *Gallup, July*, *22*.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110 (9): 2964-96.

Demir, M., Apel, R., Braga, A. A., Brunson, R. K., & Ariel, B. (2020). Body worn cameras, procedural justice, and police legitimacy: A controlled experimental evaluation of traffic stops. *Justice Quarterly*, *37*(1), 53-84.

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proceedings of the National Academy of Sciences*, *116*(34), 16793-16798.

Epp, C. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled over: How police stops define race*

and citizenship. University of Chicago Press.
Farrar, T. (2013). *Self-awareness to being watched and socially-desirable behavior: a field experiment on the effect of body-worn cameras on police use-of-force*. Police Foundation.
Froming, W. J., Walker, G. R., & Lopyan, K. J. (1982). Public and private self-awareness: When personal attitudes conflict with societal expectations. *Journal of Experimental Social Psychology*, *18*(5), 476-487.
Gaebler, J., Cai, W., Basse, G., Shroff, R., Goel, S., & Hill, J. (2020). A Causal Framework for Observational Studies of Discrimination. *arXiv preprint arXiv:2006.12460*.
Garrett, B. (2000). Standing while Black: Distinguishing" Lyons" in racial profiling cases. Columbia Law Review, 1815-1846.
Gaub, J. E., & White, M. D. (2020). Open to interpretation: Confronting the challenges of understanding the current state of body-worn camera research. *American Journal of Criminal Justice*, 1-15.
Glaser, J. (2015). *Suspect race: Causes and consequences of racial profiling*. Oxford University Press, USA.
Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254-277.
Grossmith, L., Owens, C., Finn, W., Mann, D., Davies, T., & Baika, L. (2015). Police, camera, evidence: London's cluster randomised controlled trial of body worn video. *London: College of Policing*.
Headley, A. M. (2020). Race, Ethnicity and Social Equity in Policing. In Mary E. Guy and Sean McCandless (Eds). Achieving Social Equity: From Problems to Solutions (pp. 82-91). Irvine, CA: Melvin & Leigh Publ.
Headley, A. M., & Wright, J. E. (2020). Is Representation Enough? Racial Disparities in Levels of Force and Arrests by Police. *Public Administration Review*, *80*(6), 1051-1062.
Headley, A. M., Guerette, R. T., & Shariati, A. (2017). A field experiment of the impact of body-worn cameras (BWCs) on police officer behavior and perceptions. *Journal of criminal justice*, *53*, 102-109.
Hedberg, E. C., Katz, C. M., & Choate, D. E. (2017). Body-worn cameras and citizen interactions with police officers: Estimating plausible effects given varying compliance levels. *Justice quarterly*, *34*(4), 627-651.
Houngbedji, K. (2016). Abadie's semiparametric difference-in-differences estimator. *The Stata Journal*, *16*(2), 482-490.
Huff, J. (2021). Understanding police decisions to arrest: The impact of situational, officer, and neighborhood characteristics on police discretion. *Journal of Criminal Justice*, *75*, 101829.
Knox, D., & Mummolo, J. (2020). Making inferences about racial disparities in police violence. *Proceedings of the National Academy of Sciences*, *117*(3), 1261-1262.
Kochel, T. R., Wilson, D. B., & Mastrofski, S. D. (2011). Effect of Suspect Race on Officers' Arrest Decisions. *Criminology*, *49*(2), 473-512.
Li, Y. (2022). Variation in standard errors in event-study design: insights from empirical studies and simulations. *Applied Economics*, 1-13.

Lum, C., Stoltz, M., Koper, C. S., & Scherer, J. A. (2019). Research on body-worn cameras: What we know, what we need to know. *Criminology & public policy*, *18*(1), 93-118.
Lundman, R. J., Sykes, R. E., & Clark, J. P. (1978). Police control of juveniles: A replication. *Journal of research in crime and delinquency*, *15*(1), 74-91.
MacDonald, J., Fagan, J., & Geller, A. (2016). The effects of local police surges on crime and arrests in New York City. *PLoS one*, *11*(6), e0157223.
Mastrofski, S. D., Snipes, J. B., Parks, R. B., & Maxwell, C. D. (2000). The helping hand of the law: Police control of citizens on request. *Criminology*, *38*(2), 307-342.
Mastrofski, S. D., Worden, R. E., & Snipes, J. B. (1995). Law enforcement in a time of community policing. *Criminology*, *33*(4), 539-563.
Mbuba, J. M. (2018). What if the Officer Were Black or Female? The Effects of Officer Race and Gender on Arrest Decision-Making. *African Journal of Criminology & Justice Studies*, *11*(1).
Moe, T. M. (2012). Delegation, Control, and the Study of Public Bureaucracy. *The Forum, 10*(2), Article 4. DOI: 10.1515/1540-8884.1508
Murphy, J. R. (2019). Is it recording? Racial bias, police accountability, and the body-worn camera activation policies of the ten largest metropolitan police departments in the USA. Colum. J. Race & L., 9, 141.
Neil, R., & Winship, C. (2019). Methodological challenges and opportunities in testing for racial discrimination in policing. *Annual Review of Criminology*, *2*, 73-98.
Nix, J., Todak, N., & Tregle, B. (2020). Understanding body-worn camera diffusion in US policing. *Police Quarterly*, *23*(3), 396-422.
Novak, K. J., Brown, R. A., & Frank, J. (2011). Women on patrol: An analysis of differences in officer arrest behavior. *Policing: An International Journal of Police Strategies & Management*.
Owens, C., & Finn, W. (2018). Body-worn video through the lens of a cluster randomized controlled trial in London: Implications for future research. *Policing: A Journal of Policy and Practice*, *12*(1), 77-82.
Pyo, S. (2020). Do Body-Worn Cameras Change Law Enforcement Arrest Behavior? A National Study of Local Police Departments. *The American Review of Public Administration*, 0275074020982688.
Ready, J. T., & Young, J. T. (2015). The impact of on-officer video cameras on police–citizen contacts: Findings from a controlled experiment in Mesa, AZ. *Journal of experimental criminology*, *11*(3), 445-458.
Rosenfeld, R., Johnson, T. L., & Wright, R. (2020). Are college-educated police officers different? A study of stops, searches, and arrests. *Criminal Justice Policy Review*, *31*(2), 206-236.
Sant'Anna, P. H., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, *219*(1), 101-122.
Schwartz, G. L., & Jahn, J. L. (2020). Mapping fatal police violence across US metropolitan areas: Overall rates and racial/ethnic inequities, 2013-2017. *PloS one*, *15*(6), e0229686.
Smith, D. A., & Visher, C. A. (1981). Street-level justice: Situational determinants of police arrest decisions. *Social problems*, *29*(2), 167-177.
Smith, D. A., Visher, C. A., & Davidson, L. A. (1984). Equity and discretionary justice: The influence of race on police arrest decisions. *J. Crim. L. & Criminology*, *75*, 234.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175-199.

Tregle, B., Nix, J., & Alpert, G. P. (2019). Disparity does not mean bias: Making sense of observed racial disparities in fatal officer-involved shootings with multiple benchmarks. *Journal of crime and justice*, *42*(1), 18-31.

Walker, S. (1993). *Taming the system: The control of discretion in criminal justice, 1950-1990*. Oxford University Press on Demand.

Wallace, D., White, M. D., Gaub, J. E., & Todak, N. (2018). Body-worn cameras as a potential source of depolicing: Testing for camera-induced passivity. *Criminology*, *56*(3), 481-509.

Williams Jr, M. C., Weil, N., Rasich, E. A., Ludwig, J., Chang, H., & Egrari, S. (2021). Body-Worn Cameras in Policing: Benefits and Costs.

Wright, J. E., & Headley, A. M. (2021). Can Technology Work for Policing? Citizen Perceptions of Police-Body Worn Cameras. *The American Review of Public Administration*, *51*(1), 17-27.

Yokum, D., Ravishankar, A., & Coppock, A. (2019). A randomized control trial evaluating the effects of police body-worn cameras. *Proceedings of the National Academy of Sciences*, *116*(21), 10329-10332.

7 Appendix

Table A1 Effects of BWC Adoption on Police Arrest (without pre-treatment covariates)

Dependent Variables	Coef.	S.E.	z	P>|z|	n†
Total Arrest					
Number of arrests	-195.17	97.20	-2.01	0.045**	24,400
Number of white arrests	-108.96	56.88	-1.92	0.055	24,357
Number of Black arrests	-105.78	53.51	-1.98	0.048**	21,145
Racial Disparities in Arrest					
Percentage of white arrests	0.001	0.002	0.73	0.465	24,357
Percentage of Black arrests	-0.001	0.002	-0.75	0.453	21,145
White arrests to Black arrests ratio	-0.471	1.063	-0.44	0.657	21,105
Year = 2007 to 2016, N = 33,600 (3,360 observations × 10 years).					

† Estimations were performed using observations with pairs balanced at pre-treatment and post-treatment periods.
$p < 0.05$: **, $p < 0.01$: ***, two-tailed test.